On the efficiency of code-based steganography

by

Tanjona Fiononana Ralaivaosaona

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Mathematics in the Faculty of Science at Stellenbosch University

Department of Mathematical Sciences,
Mathematics Division,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Supervisor: Prof. J.W. Sanders

March 2015

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature: .....
F.T. Ralaivaosaona

Date: February 16, 2015 .....

Abstract

On the efficiency of code-based steganography

F.T. Ralaivaosaona

Department of Mathematical Sciences,
Mathematics Division,
University of Stellenbosch,
Private Bag X1, Matieland 7602, South Africa.

Thesis: MScEng (Mech)

December 2014

Steganography is the art of hiding information inside a data host called the *cover*. The amount of distortion caused by that embedding can influence the security of the steganographic system. By secrecy we mean the detectability of the existence of the secret in the cover, by parties other than the sender and the intended recipient. Crandall (1998) proposed that coding theory (in particular the notion of covering radius) might be used to minimize embedding distortion in steganography. This thesis provides a study of that suggestion.

Firstly a method of constructing a steganographic schemes with small embedding radius is proposed by using a partition of the set of all covers into subsets indexed by the set of embeddable secrets, where embedding a secret $\mathbf{s}$ is a *maximum likelihood decoding* problem on the subset indexed by $\mathbf{s}$. This converts the problem of finding a stego-scheme with small embedding radius to a coding theoretic problem. Bounds are given on the maximum amount of information that can be embedded. That raises the question of the relationship between perfect codes and perfect steganographic schemes. We define a translation from perfect linear codes to steganographic schemes; the latter belong to the family of *matrix embedding* schemes, which arise from random linear codes. Finally, the capacity of a steganographic scheme with embedding constraint is investigated, as is the embedding efficiency to evaluate the performance of steganographic schemes.

Uittreksel

On the efficiency of code-based steganography

F.T. Ralaivaosaona

Universiteit van Stellenbosch,
Privaatsak X1, Matieland 7602, Suid Afrika.

Tesis: MScIng (Meg)

Desember 2014

Steganografie is die kuns van die wegsteek van geheime inligting in 'n data gasheer genoem die dekking. Die hoeveelheid distorsie veroorsaak deur die inbedding kan die veiligheid van die steganografiese stelsel beïnvloed. Deur geheimhouding bedoel ons die opspoorbaarheid van die bestaan van die geheim in die dekking, deur ander as die sender en die bedoelde ontvanger partye. Crandall (1998) het voorgestel dat kodeerteorie (in besonder die idee van dekking radius) kan gebruik word om inbedding distorsie te verminder in steganografie. Hierdie tesis bied 'n studie van daardie voorstel.

Eerstens 'n metode van die bou van 'n steganografiese skema met 'n klein inbedding radius word voorgestel deur die gebruik van 'n partisie van die versameling van alle dekkings in deelversamelings geïndekseer deur die versameling van inbedbare geheime, waar inbedding 'n geheime $\mathbf{s}$ is 'n *maksimum waarskynlikheid dekodering* probleem op die deelversameling geïndekseer deur $\mathbf{s}$. Dit vat die probleem van die vind van 'n stego-skema met klein inbedding radius na 'n kodering teoretiese probleem. Grense word gegee op die maksimum hoeveelheid inligting wat ingebed kan word. Dit bring op die vraag van die verhouding tussen perfekte kodes en perfekte steganographic skemas. Ons definieer 'n vertaling van perfekte lineêre kodes na steganographic skemas; laasgenoemde behoort aan die familie van matriks inbedding skemas, wat ontstaan as gevolg van ewekansige lineêre kodes. Laasten, die kapasiteit van 'n steganografiese skema met inbedding beperking word ondersoek, asook die inbedding doeltreffendheid om die prestasie van steganografiese skemas te evalueer.

Acknowledgements

First and foremost, I would like to thank God almighty who has been giving me everything to accomplish this thesis.

I would like to express my sincere gratitude to my supervisor, Prof. Jeff Sanders, who has supported me throughout my thesis with his patience, guidance and advices. Without you, this thesis would not have been completed or written.

My sincere gratitude goes also to the African Institute for Mathematical Sciences (AIMS) and the University of Stellenbosch for providing the support and funding to produce and complete my thesis.

I would like to thank all the AIMS family and stuff members: you made my life flowing and much easier. I would especially thank the following persons for their help in the writing of this thesis Jan Groenewald, Waseem Elliot and Jonathan Carter.

In my daily work I have been blessed with many friends who have been encouraging me and made my life full of happiness and laughter: I thank all of you. Ngiyabonga Siyabonga Phiwayinkosi Mthiyane: I so much appreciate your invaluable support and encouragement, which have been giving me courage and motivation to accomplish this thesis.

Last but not the least, I would like to thank my family and especially my beloved parents, Ralaivaosaona Jean Noël and Razanamampionona Jeanne d'Arc, who have been giving me encouragement, motivation, and all of support that I need in my whole life.

Dedications

Ho an'i Dada sy Mama.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

Cryptography provides the primary technique for the secure transmission of information, and has ever since the existence of the secrets. In our digital age it is no less important, in the form of public key cryptography, because of the role the web plays in supporting data transmission. But even if the canonical eavesdropper, Eve, is unable to decipher the messages passing between the canonical communicators Alice and Bob, by traffic analysis, she is able to infer more that Alice and Bob may want her to know: she observes that they are passing encrypted data.

For that reason steganography, the hiding of information (typically encrypted for safety), plays an increasingly important part. By embedding their secret in an innocuous file (like jpeg, video or audio file) Eve is unable to observe anything untoward.

The word Steganography literally means covered or hidden writing, from the Greek. Schneier *et al.* (1996) characterize steganography as a method that serves to hide secret messages in other messages, such that the secret's very existence is concealed. That is the existence of the hidden message is known only by the sender and intended receiver. Thus is also known as the science of invisible communication. A protocol which implements such a secret exchange is called steganogrphic system. It consists of

1. The cover medium(C) that will hold the secret message.
2. The secret message (M), may be plain text, digital image file or any type of data.
3. The stegonographic schemes (stego-scheme or embedding scheme), which is a couple of functions Emb and Ext for embedding and recovering the secret.

During the embedding process of a secret $\mathbf{S}$, a carrier message $\mathbf{X}$ called the *cover* is needed and then the embedding function Emb transforms the cover $\mathbf{X}$ into an innocuous looking message $\mathbf{Y}$ that must appear undistinguishable

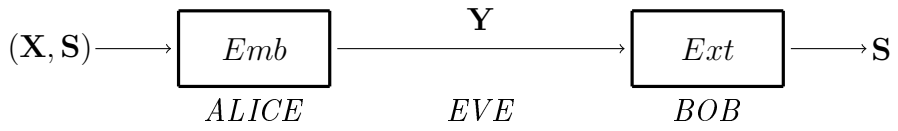

Figure 1.1: A model of steganographic system.

from $\mathbf{X}$. The output $\mathbf{Y} = Emb(\mathbf{X}, \mathbf{S})$ is called the *stego*. A stego-key might be needed to hide and recover the secret. It might be used to increase the security of the steganographic system, by combining steganography with cryptography.

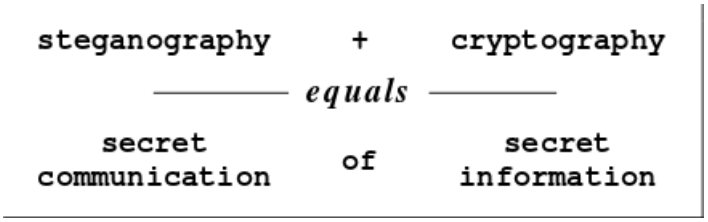

Figure 1.2: Steganography and Cryptography (Engle, 2003)

The main requirement of steganography, undetectability, means that an attacker is able to distinguish between stego and cover objects with success no better than random guessing, given the knowledge of the embedding function and the source of the cover media. The formal definition of secure steganographic system was given by Cachin (1998). It is based on minimizing the success probability of the adversary to guess whether or not a message is steganographic. Detectability is influenced by many factors, such as the type of the cover object, the selection of the places that could be modified during embedding, the embedding operations, and the number of changes caused by the embedding operation. If two embedding methods share the first three factors, then the one that introduces fewer changes will (typically) be less detectable.

The most important problem in steganogarphy is formalizing that very concept. Any formalization should provide a foundation for analysing detectability of an embedded secret in a cover-text. It seems inevitable that information theory be used. Cachin (1998) formalised detectability using relative entropy (which has also been the concern of my previous work (Ralaivaosaona, May 2013)). Zöllner *et al.* (1998) used Shannon's mutual information. Many other authors have studied this concept but they use one of the above approaches, but mostly Cachin's.

The next important problems concern establishing bounds and results on the efficiency, which is also has to be formalised. That includes the evaluation of the hiding capacity which upper-bounds the rates of embeddable information

and the fundamental trade-off between the achievable rates and the allowed distortion. Ker *et al.* (2008) has established some results on the steganographic capacity such as the square root law. Moulin and O'Sullivan (2003) introduced a more general result on the hiding capacity of any information hiding case with embedding constraint, which is a distortion parameter D. The strength of the transparency constraint is controlled by that distortion parameter which should (in general) be small, as embedding is intended to be imperceptible.

We follow the view that (in general) the fewer changes needed to embed a secret in a cover-text, the lower the detectability. In this thesis coding theory provides a model of embedding with minimum distortion.

Westfeld (2001) introduced the concept of *embedding efficiency*, which is the expected number of random bits embedded per one embedding change. A good scheme must have as high on embedding efficiency as possible. In 1998, Crandall (1998) showed that embedding efficiency of steganographic schemes can be improved by applying covering codes to the embedding process. In particular, linear codes can be used to construct an embedding scheme whose embedding capacity is the code redundancy, and the covering radius of the code corresponds to the maximum number of embedding changes necessary to embed any message. From then many authors have developed a theory connecting coding theory and steganography. Galand and Kabatiansky (2003*b*) and Munuera (2012) gave an explicit connection between a collection of codes and a steganographic scheme, Zhang and Li (2005) introduced the notion of steganographic codes and explored the connection between maximum length embeddable (MLE) steganographic codes and perfect codes, and Westfeld (2001) implemented Crandall's (Crandall, 1998) idea of F5 steganography.

The F5 algorithm is a practical method of embedding bits in digital images, more precisely on the least significant bits (LSB) of its pixel values. This method has been known to resist statistical attack. It has reasonable efficiency since it is capable of embedding k information bits in a sequence of 2^k-1 pixels (LSB's) by flipping at most 1 pixel value. That is because it uses uses Hamming codes to embed data, and this family of codes has redundancy k, length 2^k-1 and covering radius 1. In term of coding theory, it is a very important family of codes known as *perfect* codes, which can correct all errors up to the covering radius. Hamming codes are single error correcting codes.

In this thesis, we focus on the construction of "good" stego-schemes. "Good" in terms of minimizing the embedding impact in order to increase embedding efficiency. Some ideas of Munuera (2012) will be used and extended in the way I understand them. So the first chapter is the formalisation of a better scheme together with its construction. Then we will restrict that idea to a coding theoretic method. It will become more interesting since we give some bounds on the performance of the constructed stego-scheme from any code (code-based steganography). Most of those bounds are derived from coding theoretic bounds like Hamming bound, covering bound, etc. A specialization of our construction is Crandall's (Crandall, 1998) matrix encoding. In this

case, stego-schemes are constructed from linear codes. Linear codes have better properties, therefore, we can easily express and compute the parameters of matrix encoding (a scheme derived from random linear code) with respect to the parameters of the code. We will see that some bounds on the performances of code-based stego-schemes can be achieved by random linear codes.

Chapter 2

Construction of good steganographic schemes

Steganography is the method of hiding secret messages inside a cover-object[1]. To communicate a secret covertly, Alice and Bob may proceed in three different ways.

- *cover selection* where the sender selects an appropriate cover-object that will communicate the desired message. The choice of the cover-object depends on the message to be hidden;
- *cover synthesis* in the case the embedder has to generate the cover-object from the message to be hidden;
- *cover modification* is used most frequently, where the embedder has some large source of cover-objects and he embeds the message into an arbitrary one by modifying some parts of it.

This thesis only focuses on embedding by cover modification, where Alice chooses a cover-object and then modifies it (or part of it) in order to convey the desired secret in a manner such that it is hidden. That is, after embedding has taken place, the original and the altered cover-object or *stego-object*[2] must be seemingly identical so that no one apart from Alice and the receiver, Bob, could be able to tell whether the transmitted message carries hidden information or not. This means that stego and cover have to be statistically indistinguishable, where statistical detectability of most steganographic schemes increases with *embedding distortion* (Fridrich *et al.*, 2007*b*). Therefore it is important for Alice to embed the secret while introducing as small an impact to the cover as possible.

[1] A cover-object consists a non specific carrier data. It can be an image, a text, or sequence of symbols,... Sometimes we only use "cover" for simplicity.

[2] "Stego-object" refers to a message or object that contains secret information.

Throughout this thesis, secrets and cover (and stego) are represented by boldface symbols. This is to differentiate a symbol and a sequence of symbols in the case where covers and/or secrets are sequences from an alphabet, that we can consider as vectors. So boldface symbols also stand for vectors or matrices. Calligraphic font is used for alphabets (or sets).

This chapter gives an overall idea of how to construct steganographic schemes with small distortion and it is organised as follows. In Section 2.1, basic definitions and notation on steganography are introduced. We give some useful properties in Section 2.2. Then, Section 2.3 focuses on the constructions of "good" schemes. We close the chapter by introducing a partial order on the space of all embedding schemes defined on the same set of covers and secrets.

2.1 Steganographic scheme (Stego-scheme)

We assume that $\mathcal{M}$ is the set of all embeddable messages (or secrets), and $\mathcal{X}$ the set of all cover-objects, such that[3] $|\mathcal{M}|$ and $|\mathcal{X}|$ are finite with $|\mathcal{M}| < |\mathcal{X}|$. Then we can define a stego-scheme as follows:

Definition 2.1. *A* stego-scheme $\mathcal{S} = (Emb, Ext; \mathcal{X}, \mathcal{M})$ *is a pair of Embedding, Emb, and Extracting, Ext, functions defined between* $\mathcal{X}$ *and* $\mathcal{M}$*:*

$$\begin{aligned} Emb &: \mathcal{X} \times \mathcal{M} \to \mathcal{X} \\ Ext &: \mathcal{X} \to \mathcal{M} \end{aligned}$$

such that for all $\mathbf{x} \in \mathcal{X}$ *and* $\mathbf{s} \in \mathcal{M}$*,*

$$Ext(Emb(\mathbf{x}, \mathbf{s})) = \mathbf{s}. \tag{2.1.1}$$

$\mathbf{y} = Emb(\mathbf{x}, \mathbf{s})$ *is called the stego-object.*

The embedding function Emb takes the cover-object $\mathbf{x}$ and the secret $\mathbf{s}$ as its inputs and produces the stego-object in such a way that we can always recover the secret from the resulting stego-object using the extracting function Ext.

For a given embedding function, the cardinality, $|\mathcal{M}|$, of the set $\mathcal{M}$, is the number of different messages that can be communicated. The logarithm $\log_2 |\mathcal{M}|$, is called *embedding capacity*. Its unit is in bits and it is denoted by h.

Defining a distance $d : \mathcal{X} \times \mathcal{X} \to [0, +\infty)$, we measure the impact of embedding (or *embedding distortion*) as follows.

Definition 2.2. *Let* $\mathcal{S} = (Emb, Ext; \mathcal{X}, \mathcal{M})$ *be a stego-scheme. The* embedding distortion *introduced to a cover* $\mathbf{x} \in \mathcal{X}$ *by the function Emb in order to hide the secret* $\mathbf{s} \in \mathcal{M}$ *is given by* $d(\mathbf{x}, Emb(\mathbf{x}, \mathbf{s}))$*.*

[3] $|E|$ stands for the cardinality of the set E.

We denote the expected embedding distortion taken over uniformly distributed secrets and covers by R_a, i.e.

$$R_a = E(d(\mathbf{x}, Emb(\mathbf{x}, \mathbf{s}))). \tag{2.1.2}$$

The worst-case embedding distortion is given by the *embedding radius*. It gives the maximum possible distortion that can be made, over all possible covers and secrets.

Definition 2.3. *Let $\mathcal{S} = (Emb, Ext; \mathcal{X}, \mathcal{M})$ be a stego-scheme. The* embedding radius *R of $\mathcal{S}$ is given by*

$$R := \max\{d(\mathbf{x}, Emb(\mathbf{x}, \mathbf{s}))|\mathbf{x} \in \mathcal{X}, \mathbf{s} \in \mathcal{M}\}. \tag{2.1.3}$$

Obviously we have $R_a \leq R$.

The efficiency of a stego-scheme is usually evaluated through its *embedding efficiency*, which is defined as follows.

Definition 2.4. *(Westfeld, 2001) The* embedding efficiency*, e, is the expected number of embedded bits per unit distortion. It is given by the ratio between the embedding capacity and the expected embedding distortion, i.e.*

$$e := \frac{h}{R_a}. \tag{2.1.4}$$

Similarly the lower embedding efficiency $\underline{e}$ is the ratio between embedding capacity and embedding radius, that is

$$\underline{e} := \frac{h}{R}. \tag{2.1.5}$$

Since $R \geq R_a$, it follows that $\underline{e} \leq e$.

We can compare any two stego-schemes defined on the same set of covers and secrets by their embedding efficiencies. The one with smaller embedding distortion is less detectable than the other. Similarly, the one with higher embedding efficiency is better than the other. So given the two sets $\mathcal{X}$ and $\mathcal{M}$, we aim to design stego-schemes with maximal embedding efficiency. Equivalently, we aim to minimize embedding distortion.

2.2 Property of stego-schemes

From Definition 2.1, we can derive the following properties on both embedding and extracting function; Emb and Ext, given that the cover set is $\mathcal{X}$ and the set of embeddable secrets is $\mathcal{M}$.

Proposition 2.5. *Let $\mathcal{S} = (Emb, Ext; \mathcal{X}, \mathcal{M})$ be a stego-scheme. Then*

1. *$Ext : \mathcal{X} \to \mathcal{M}$ is a surjective function;*

2. *for each $\mathbf{x} \in \mathcal{X}$, the map $Emb(\mathbf{x}, .) : \mathcal{M} \to \mathcal{X}$ is injective.*

Proof. The proof of both statements follows easily from Equation 2.1.1. □

For any $\mathbf{s} \in \mathcal{M}$, we define the inverse image of the singleton $\{\mathbf{s}\} \subset \mathcal{M}$ as the set

$$Ext^{-1}(\{\mathbf{s}\}) := \{\mathbf{y} \in \mathcal{X} | Ext(\mathbf{y}) = \mathbf{s}\}.$$

Since Ext is a surjective function, the union of all inverse images of each singleton $\{\mathbf{s}\} \subset \mathcal{M}$ of all elements of $\mathcal{M}$ cover the whole space $\mathcal{X}$. Moreover, it partitions $\mathcal{X}$ since any two inverse images of two different singletons $\{\mathbf{s}\}, \{\mathbf{s}'\}$ are disjoint. That is, for any distinct $\mathbf{s}, \mathbf{s}' \in \mathcal{M}$,

$$Ext^{-1}(\{\mathbf{s}\}) \cap Ext^{-1}(\{\mathbf{s}'\}) = \emptyset.$$

Then we can derive the following equivalence.

Proposition 2.6. *The following are equivalent:*

1. *An extracting function $Ext : \mathcal{X} \to \mathcal{M}$ that enables us to extract, from any cover-object $\mathbf{x} \in \mathcal{X}$, a secret $\mathbf{s} \in \mathcal{M}$.*

2. *An $|\mathcal{M}|$-partition[4] (pointed) of $\mathcal{X}$, indexed by elements of $\mathcal{M}$.*

Proof. The surjectivity of Ext implies that[5] $\mathcal{X} = \sqcup_{\mathbf{s} \in \mathcal{M}} Ext^{-1}(\{\mathbf{s}\})$ and each set $Ext^{-1}(\{\mathbf{s}\})$ contains at least one element. Moreover, for any two different secrets $\mathbf{s}$ and $\mathbf{s}'$, $Ext^{-1}(\{\mathbf{s}\}) \cap Ext^{-1}(\{\mathbf{s}'\}) = \emptyset$. Otherwise if $\mathbf{y} \in Ext^{-1}(\{\mathbf{s}\}) \cap Ext^{-1}(\{\mathbf{s}'\})$, then $Ext(\mathbf{y}) = \mathbf{s}$ and $Ext(\mathbf{y}) = \mathbf{s}'$, which is impossible since Ext is a function and $\mathbf{s} \neq \mathbf{s}'$.

Conversely, if we have an $|\mathcal{M}|$-partition of $\mathcal{X}$, indexed by elements of $\mathcal{M}$, say $\{\mathcal{X}_{\mathbf{s}} \,|\, \mathbf{s} \in \mathcal{M}\}$, then define an extracting function $Ext : \mathcal{X} \to \mathcal{M}$ such that $Ext(\mathbf{x}) = \mathbf{s}$ if $\mathbf{x} \in \mathcal{X}_{\mathbf{s}}$. □

If we have an extracting function, $Ext : \mathcal{X} \to \mathcal{M}$ (equivalently, an $|\mathcal{M}|$-partition of $\mathcal{X}$), then for any embedding function Emb we could choose, the Relation 2.1.1 must hold for Ext and Emb to form a stego-scheme. An example is given below.

Example 2.7. *Let Ext be an extracting function and $\mathbf{s} \in \mathcal{M}$ be a secret. The embedding function Emb* randomly *modifies any element of $\mathcal{X}$ to an element of $Ext^{-1}(\{\mathbf{s}\})$.*

[4] An M-partition of any set $\mathcal{X}$ is defined for any integer $M \geq 1$ as a partition of $\mathcal{X}$ containing M subsets (not considering the empty set).

[5] $\sqcup$ is disjoint union.

It is obvious by construction that $(Emb, Ext; \mathcal{X}, \mathcal{M})$ is a stego-scheme since Equation 2.1.1 holds. But random modification does not guarantee that cover and stego are close enough with respect to a distance d on $\mathcal{X}$. With probability $\frac{1}{|Ext^{-1}(\{\mathbf{s}\})|}$, any cover $\mathbf{x}$ is modified to the most distant element of $Ext^{-1}(\{\mathbf{s}\})$, i.e.

$$d(\mathbf{x}, Emb(\mathbf{x}, \mathbf{s})) = \max_{\mathbf{y} \in Ext^{-1}(\{\mathbf{s}\})} d(\mathbf{x}, \mathbf{y}).$$

Munuera (2012) said that any stego-scheme $\mathcal{S} = (Emb, Ext; \mathcal{X}, \mathcal{M})$ can be refined to give a better scheme with smaller embedding distortion and it can be done by Algorithm 2.1.

Algorithm 2.1 Refinement of the embedding function Emb.

1. Check if there are $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ and $\mathbf{s} \in \mathcal{M}$ such that $Ext(\mathbf{y}) = \mathbf{s}$ and $d(\mathbf{x}, \mathbf{y}) < d(\mathbf{x}, Emb(\mathbf{x}, \mathbf{s}))$.

2. If such $\mathbf{x}, \mathbf{y}, \mathbf{s}$ exist, then define Emb' such that

$$\begin{cases} Emb'(\mathbf{x}, \mathbf{s}) = \mathbf{y} \\ Emb'(\mathbf{x}', \mathbf{s}') = Emb(\mathbf{x}', \mathbf{s}') \quad \text{for all } (\mathbf{x}, \mathbf{s}) \neq (\mathbf{x}', \mathbf{s}'). \end{cases}$$

Note that for all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{s} \in \mathcal{M}$, Emb' decreases embedding distortion:

$$d(\mathbf{x}, Emb'(\mathbf{x}, \mathbf{s})) \leq d(\mathbf{x}, Emb(\mathbf{x}, \mathbf{s})). \tag{2.2.1}$$

For fixed extracting function $Ext : \mathcal{X} \to \mathcal{M}$, let $\mathbb{E}$ be the set of all embedding functions, Emb, such that $(Emb, Ext; \mathcal{X}, \mathcal{M})$ is a stego-scheme. We define on $\mathbb{E}$ the refinement relation $\sqsubseteq$, such that: $Emb \sqsubseteq Emb'$ if and only if Emb' is defined from Emb according to Algorithm 2. Therefore we have

$$Emb \sqsubseteq Emb' \Rightarrow d(\mathbf{x}, Emb'(\mathbf{x}, \mathbf{s})) \leq d(\mathbf{x}, Emb(\mathbf{x}, \mathbf{s})). \tag{2.2.2}$$

After finitely many consecutive (say N) refinement steps of the embedding function, Emb, we arrive at a point were any modification is no longer an improvement on the embedding distortion. Then denote the output of the final step as Emb^*. That is

$$Emb \sqsubseteq Emb' \sqsubseteq \cdots \sqsubseteq Emb^*, \tag{2.2.3}$$

and therefore we have

$$d(\mathbf{x}, Emb^*(\mathbf{x}, \mathbf{s})) \leq d(\mathbf{x}, Emb(\mathbf{x}, \mathbf{s})) \tag{2.2.4}$$

for all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{s} \in \mathcal{M}$. Actually, that is the best we can do to improve embedding distortion. The stego-scheme $\mathcal{S}^* = (Emb^*, Ext; \mathcal{X}, \mathcal{M})$ is *proper* if and only if Algorithm 2.1 is no longer applicable. A definition of *proper* embedding scheme is given as follows.

Definition 2.8. *(Munuera, 2012) Let* $\mathcal{S} = (Emb, Ext; \mathcal{X}, \mathcal{M})$ *be a stego-scheme. Then* $\mathcal{S}$ *is proper if the embedding distortion is the minimum allowed by* Ext*. That is for all* $\mathbf{x} \in \mathcal{X}$*,* $\mathbf{s} \in \mathcal{M}$ *and a distance* d *on* $\mathcal{X}$*,*

$$d(\mathbf{x}, Emb(\mathbf{x}, \mathbf{s})) = d(\mathbf{x}, Ext^{-1}(\{\mathbf{s}\})) = \min_{\mathbf{y} \in Ext^{-1}(\{\mathbf{s}\})} d(\mathbf{x}, \mathbf{y}). \tag{2.2.5}$$

Let us adopt the refinement relation $\sqsubseteq$ on the stego-scheme itself, such that $(Emb, Ext; \mathcal{X}, \mathcal{M}) \sqsubseteq (Emb', Ext; \mathcal{X}, \mathcal{M}) \Leftrightarrow Emb \sqsubseteq Emb'$.

Proposition 2.9. *If* $\mathcal{S} = (Emb, Ext; \mathcal{X}, \mathcal{M}) \sqsubseteq \mathcal{S}^* = (Emb^*, Ext; \mathcal{X}, \mathcal{M})$ *and if* $\mathcal{S}$ *and* $\mathcal{S}^*$ *have embedding efficiencies* e *and* e^* *respectively, then*

$$e \leq e^*.$$

Similarly, if the lower embedding efficiencies are respectively $\underline{e}$ *and* $\underline{e}^*$ *then*

$$\underline{e} \leq \underline{e}^*.$$

Proof. The proof of the proposition follows easily from Equation 2.2.4. □

Since proper schemes have better parameters and any non-proper[6] embedding scheme can be modified to become proper according to Algorithm 2.1, then from now on we consider only embedding functions that are proper. The next section focuses on the construction to design this kind of scheme.

2.3 Construction of good schemes

This construction focuses not only on the embedding function but on the extraction function as well. They both influence the quality of the stego-scheme. Firstly, assume the extracting function, $Ext : \mathcal{X} \to \mathcal{M}$, is arbitrary; we construct a suitable embedding function $Emb : \mathcal{X} \times \mathcal{M} \to \mathcal{X}$ such that $(Emb, Ext; \mathcal{X}, \mathcal{M})$ is proper.

2.3.1 Proper stego-schemes

For a given extracting function Ext, the best strategy to embed a secret $\mathbf{s} \in \mathcal{M}$ inside a cover $\mathbf{x} \in \mathcal{X}$ with minimum distortion (with respect to Ext) is to find the closest $\mathbf{y} \in Ext^{-1}(\{\mathbf{s}\})$ to the cover $\mathbf{x} \in \mathcal{X}$. That method is called the *Maximum-likelihood decoding* problem in (Barbier, 2010).

Definition 2.10. *(Barbier, 2010) Let* $\mathcal{C}$ *be a subset of* $\mathcal{X}$ *and* $\mathbf{x} \in \mathcal{X}$*. The* maximum-likelihood decoding *finds an element* $\mathbf{y} \in \mathcal{C}$ *closest to* $\mathbf{x}$*. More precisely, it finds* $\mathbf{y} \in \mathcal{C}$*, such that*

$$d(\mathbf{x}, \mathbf{y}) = d(\mathbf{x}, \mathcal{C}) := \min_{\mathbf{c} \in \mathcal{C}} d(\mathbf{x}, \mathbf{c}). \tag{2.3.1}$$

[6] A scheme is non-proper if Algorithm 2.1 is still applicable.

We define the relation $Dec_{\mathcal{C}} : \mathcal{X} \leftrightarrow \mathcal{C}$ such that for all $\mathbf{x} \in \mathcal{X}$,

$$Dec_{\mathcal{C}}(\{\mathbf{x}\}) = \{\mathbf{y} \in \mathcal{C} | d(\mathbf{x}, \mathbf{y}) = d(\mathbf{x}, \mathcal{C})\}. \tag{2.3.2}$$

Proposition 2.11. *If we define $\mathcal{S} = (Emb, Ext; \mathcal{X}, \mathcal{M})$ such that for all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{s} \in \mathcal{M}$,*

$$Emb(\mathbf{x}, \mathbf{s}) \in Dec_{Ext^{-1}(\{\mathbf{s}\})}(\{\mathbf{x}\}), \tag{2.3.3}$$

then $\mathcal{S}$ is a proper stego-scheme.

Proof. Since $Emb(\mathbf{x}, \mathbf{s}) \in Ext^{-1}(\{\mathbf{s}\})$, therefore $Ext(Emb(\mathbf{x}, \mathbf{s})) = \mathbf{s}$. That is $\mathcal{S}$ is a stego-scheme. Moreover, $Emb(\mathbf{x}, \mathbf{s}) \in Dec_{Ext^{-1}(\{\mathbf{s}\})}(\{\mathbf{x}\})$, i.e.

$$d(\mathbf{x}, Emb(\mathbf{x}, \mathbf{s})) = d(\mathbf{x}, Ext^{-1}(\{\mathbf{s}\}))$$

and therefore $\mathcal{S}$ is proper by Definition 2.8. □

Let Ext be an extracting function. The set $\{\mathcal{X}_{\mathbf{s}} | \mathbf{s} \in \mathcal{M}\}$, such that $\mathcal{X}_{\mathbf{s}} = Ext^{-1}(\{\mathbf{s}\})$, is an $|\mathcal{M}|$-partition of $\mathcal{X}$ (see Proposition 2.6). So by Proposition 2.11, the stego-scheme $(Emb, Ext; \mathcal{X}, \mathcal{M})$ such that for all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{s} \in \mathcal{M}$,

$$Emb(\mathbf{x}, \mathbf{s}) \in Dec_{\mathcal{X}_{\mathbf{s}}}(\mathbf{x}), \tag{2.3.4}$$

is a proper stego-scheme with respect to Ext. Therefore it has the maximum embedding efficiency among all schemes[7] $(., Ext, \mathcal{X}, \mathcal{M})$ according to Proposition 2.9.

This method is then efficient in terms of minimizing embedding distortion for any cover and secret.

2.3.2 T-Covering

Here we impose a threshold T on the embedding distortion with respect to a distance d on $\mathcal{X}$.

Definition 2.12. *(Galand and Kabatiansky, 2003a) An embedding scheme with quality threshold T is a pair of functions, an embedding function $Emb : \mathcal{X} \times \mathcal{M} \to \mathcal{X}$ and an extracting function $Ext : \mathcal{X} \to \mathcal{M}$, such that for any $\mathbf{x} \in \mathcal{X}$ and for any $\mathbf{s} \in \mathcal{M}$, the stego-object $\mathbf{y} = Emb(\mathbf{x}, \mathbf{s}) \in \mathcal{X}$ must satisfy*

1. $Ext(\mathbf{y}) = \mathbf{s}$,
2. [8] $d(\mathbf{x}, \mathbf{y}) \leq T$.

[7] $(., Ext, \mathcal{X}, \mathcal{M})$ is the set of all stego-schemes that share the same extracting function Ext.

[8] $d(\mathbf{x}, \mathbf{y}) \leq T$ means that the covering radius $R = T$.

This definition means that for any cover $\mathbf{x}$ and any secret $\mathbf{s}$, we can embed $\mathbf{s}$ in $\mathbf{x}$ with embedding distortion not more than T. We can construct such a scheme by using an $|\mathcal{M}|$-partition of $\mathcal{X}$ with *covering radius*[9] T.

Definition 2.13. *Let $M > 0$ be an integer and $\mathcal{P}_M = \{\mathcal{X}_i | i \in [1, M]\}$ be an M-partition of $\mathcal{X}$. Then we define the* covering radius *of $\mathcal{P}_M$ to be the smallest integer T such that*[10]

$$d(\mathbf{x}, \mathcal{X}_i) \leq T.$$

*An M-partition of $\mathcal{X}$ with covering radius T is called an (M, T)-*covering *of $\mathcal{X}$ or simply T-*covering.

The following proposition shows how to use coverings of $\mathcal{X}$ to embed data.

Proposition 2.14. *Let $\{\mathcal{X}_\mathbf{s} | \mathbf{s} \in \mathcal{M}\}$ be an $(|\mathcal{M}|, T)$-covering of $\mathcal{X}$. Consider the functions $Ext : \mathcal{X} \to \mathcal{M}$ and $Emb : \mathcal{X} \times \mathcal{M} \to \mathcal{X}$ defined by $Ext(\mathbf{x}) = \mathbf{s}$ if $\mathbf{x} \in \mathcal{X}_\mathbf{s}$ and $Emb(\mathbf{x}, \mathbf{s}) \in Dec_{\mathcal{X}_\mathbf{s}}(\mathbf{x})$. Then $\mathcal{S} = (Emb, Ext; \mathcal{X}, \mathcal{M})$ is a stego-scheme with quality threshold T and it is proper.*

Proof. Since $Emb(\mathbf{x}, \mathbf{s}) \in Dec_{\mathcal{X}_\mathbf{s}}$, we have $Ext(Emb(\mathbf{x}, \mathbf{s})) = \mathbf{s}$. Furthermore, $d(\mathbf{x}, Emb(\mathbf{x}, \mathbf{s})) = d(\mathbf{x}, \mathcal{X}_\mathbf{s}) \leq T$ by definition of $(|\mathcal{M}|, T)$-covering of $\mathcal{X}$ (see Definition 2.13). By construction $\mathbf{S}$ is proper (see Proposition 2.11). □

The converse also holds. We can define from a stego-scheme with quality threshold T an $(|\mathcal{M}|, T)$-covering of $\mathcal{X}$, but that is not our issue. It is a new problem in coding theory called *Steganographic code* by Zhang and Li (2005).

Proposition 2.15. *The following are equivalent:*

1. *A proper stego-scheme $\mathcal{S} = (Emb, Ext; \mathcal{X}, \mathcal{M})$ with quality threshold T.*
2. *An $(|\mathcal{M}|, T)$-covering of $\mathcal{X}$.*

Proof. From $(|\mathcal{M}|, T)$-covering to stego-scheme, we use Proposition 2.14. Conversely, for all $\mathbf{s} \in \mathcal{M}$, the set $\mathcal{P} = \{Ext^{-1}(\{\mathbf{s}\}) | \mathbf{s} \in \mathcal{M}\}$ is an $|\mathcal{M}|$-partition of $\mathcal{X}$ and for all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{s} \in \mathcal{M}$, $d(\mathbf{x}, Ext^{-1}(\{\mathbf{s}\})) \leq T$. □

2.4 Ordering stego-schemes

Let us define the space $(\mathbb{S}, \preccurlyeq_\mathbb{S})$ such that $\mathbb{S}$ is the set of all stego-schemes defined between $\mathcal{X}$ and $\mathcal{M}$ and the relation $\preccurlyeq_\mathbb{S}$ is defined such that: if $\mathcal{S} = (Emb, Ext; \mathcal{X}, \mathcal{M}), \mathcal{S}' = (Emb', Ext'; \mathcal{X}, \mathcal{M}) \in \mathbb{S}$, then

$$\mathcal{S} \preccurlyeq_\mathbb{S} \mathcal{S}' \iff R \geq R', \tag{2.4.1}$$

[9] Covering radius is a term used in coding theory but we adopt it here to express our problem clearly.

[10] Given a distance d on $\mathcal{X}$, we define the distance between a subset $\mathcal{C}$ and an element $\mathbf{x}$ of $\mathcal{X}$ to be $d(\mathbf{x}, \mathcal{C}) = \min_{\mathbf{y} \in \mathcal{C}} d(\mathbf{x}, \mathbf{y})$.

where R and R' are respectively the embedding radii of $\mathcal{S}$ and $\mathcal{S}'$, and it means that $\mathcal{S}'$ is better than $\mathcal{S}$. If we assume that any two schemes having the same embedding radius are equivalent, then $(\mathbb{S}, \preccurlyeq_{\mathbb{S}})$ is a *partially-ordered space.*

There is a relationship between $\sqsubseteq$ (see Section 2.2) and $\preccurlyeq_{\mathbb{S}}$, and it is given as follows.

Proposition 2.16. *Let $\mathcal{S} = (Emb, Ext; \mathcal{X}, \mathcal{M}) \in \mathbb{S}$ be a non-proper stego-scheme, and $\mathcal{S}' = (Emb', Ext; \mathcal{X}, \mathcal{M}) \in \mathbb{S}$ such that Emb' is a refinement of Emb. Then we have*

$$\mathcal{S} \sqsubseteq \mathcal{S}' \Longrightarrow \mathcal{S} \preccurlyeq_{\mathbb{S}} \mathcal{S}'.$$

Proof. If $\mathcal{S} \sqsubseteq \mathcal{S}'$, then for all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{s} \in \mathcal{M}$,

$$d(\mathbf{x}, Emb'(\mathbf{x}, \mathbf{s})) \leq d(\mathbf{x}, Emb(\mathbf{x}, \mathbf{s})).$$

By the definition of embedding radius (see Definition 2.3), $R \geq R'$. Thus $\mathcal{S} \preccurlyeq_{\mathbb{S}} \mathcal{S}'$. □

We define the set $\mathbb{P}$ of all $|\mathcal{M}|$-partitions of $\mathcal{X}$. Let $\preccurlyeq_{\mathbb{P}}$ be a relation defined on $\mathbb{P}$ such that: if $\mathcal{P}_{\mathcal{X}} = \{\mathcal{X}_1, \ldots, \mathcal{X}_{|\mathcal{M}|}\}$ is a ρ-covering[11] of $\mathcal{X}$ and $\mathcal{P'}_{\mathcal{X}} = \{\mathcal{X'}_1, \ldots, \mathcal{X'}_{|\mathcal{M}|}\}$ is a ρ'-covering of $\mathcal{X}$, then

$$\mathcal{P}_{\mathcal{X}} \preccurlyeq_{\mathbb{P}} \mathcal{P'}_{\mathcal{X}} \Longleftrightarrow \rho \geq \rho'. \tag{2.4.2}$$

Evidently, $\preccurlyeq_{\mathbb{P}}$ is a pre-order. As usual it extends to a partial order on $\preccurlyeq_{\mathbb{P}}$-equivalence classes.

Let τ define the transformation, from covering of $\mathcal{X}$ to stego-schemes, given in Proposition 2.14. That is if $\mathcal{P}_{\mathcal{X}} = \{\mathcal{X}_{\mathbf{s}} | \mathbf{s} \in \mathcal{M}\}$ is a ρ-covering of $\mathcal{X}$, then define

$$\tau(\mathcal{P}_{\mathcal{X}}) := (Emb, Ext; \mathcal{X}, \mathcal{M}), \tag{2.4.3}$$

such that $Ext(\mathbf{x}) = \mathbf{s}$ if $\mathbf{x} \in \mathcal{X}_{\mathbf{s}}$ and $Emb(\mathbf{x}, \mathbf{s}) \in Dec_{\mathcal{X}_{\mathbf{s}}}(\mathbf{x})$.

Proposition 2.17. *τ is isotone, i.e. if $\mathcal{P}_{\mathcal{X}} \preccurlyeq_{\mathbb{P}} \mathcal{P'}_{\mathcal{X}}$, then $\tau(\mathcal{P}_{\mathcal{X}}) \preccurlyeq_{\mathbb{S}} \tau(\mathcal{P'}_{\mathcal{X}})$. Moreover, if $\mathcal{P}_{\mathcal{X}}$ is a ρ-covering of $\mathcal{X}$ (resp. $\mathcal{P'}_{\mathcal{X}}$ is a ρ'-covering), then $\tau(\mathcal{P}_{\mathcal{X}})$ has embedding radius $R = \rho$ (resp. $\tau(\mathcal{P'}_{\mathcal{X}})$ has embedding radius $R' = \rho'$).*

Proof. Let $\mathcal{P}_{\mathcal{X}} = \{\mathcal{X}_{\mathbf{s}} | \mathbf{s} \in \mathcal{M}\}$ and $\mathcal{P'}_{\mathcal{X}} = \{\mathcal{X'}_{\mathbf{s}} | \mathbf{s} \in \mathcal{M}\}$ and let ρ and ρ' be their covering radii respectively. If we assume that

$$\tau(\mathcal{P}_{\mathcal{X}}) = (Emb, Ext; \mathcal{X}, \mathcal{M}) = \mathcal{S},$$

$$\tau(\mathcal{P'}_{\mathcal{X}}) = (Emb', Ext'; \mathcal{X}, \mathcal{M}) = \mathcal{S}'$$

[11] Partition of $\mathcal{X}$ with covering radius ρ.

and they respectively have embedding radii R and R', then we have

$$\begin{aligned}
\mathcal{P}_{\mathcal{X}} \preccurlyeq_{\mathbb{P}} \mathcal{P}'_{\mathcal{X}} &\iff \rho \geq \rho' \\
&\iff \forall \mathbf{x}, \mathbf{s} : d(\mathbf{x}, \mathcal{X}_{\mathbf{s}}) \geq d(\mathbf{x}, \mathcal{X}'_{\mathbf{s}}) \\
&\iff \max_{\mathbf{x},\mathbf{s}} d(\mathbf{x}, \mathcal{X}_{\mathbf{s}}) \geq \max_{\mathbf{x},\mathbf{s}} d(\mathbf{x}, \mathcal{X}'_{\mathbf{s}}) \\
&\iff R \geq R' \\
&\iff \mathcal{S} \preccurlyeq_{\mathbb{S}} \mathcal{S}'.
\end{aligned}$$

□

A "good" stego-scheme $\mathcal{S} = (Emb, Ext; \mathcal{X}, \mathcal{M}) \in \mathbb{S}$ derives from an $|\mathcal{M}|$-partition of $\mathcal{X}$ with the smallest covering radius possible with respect to a suitable distance d defined on $\mathcal{X}$. Moreover, there is no other stego-scheme $\mathcal{S}' \in \mathbb{S}$ such that $\mathcal{S} \sqsubseteq \mathcal{S}'$. That means, $\mathcal{S}$ is proper. An explicit example of a scheme with the smallest embedding radius is given in Chapter 4, where covers and secrets are represented by bit strings.

Chapter 3

Application of coding theory to steganography: Code-based steganography

In this chapter we use sequences from an alphabet $\mathcal{A}$ to be the covers, called *cover-sequences* (or -word or -text). The set of secrets is $\mathcal{M}$. The resulting stego is also a sequence from $\mathcal{A}$ with the same length as the cover, called the *stego-sequence* (or -word or -text). The distortion is then the number of changes introduced by the embedding function in the cover. That number of changes is captured by the Hamming distance (See Definition 3.6) between the two equi-length sequences: the cover and the stego-sequence. We consider Hamming distance because it is the metric used in coding theory. Since we are linking the two areas, it is fundamentally relevant.

Let the alphabet $\mathcal{A} = \{a_1, a_1, \ldots, a_q\}$. For example, in steganography using 8-bit gray-scale digital images, $\mathcal{A}$ is the set of all integers in the range of $[0, 256)$. If we define a code as just a subset of $\mathcal{A}^n$, then the construction in the previous chapter needs to handle $|\mathcal{M}|$ different codes, each with their cardinality and with its own decoding strategy. It would be easier if the subsets that partition the space $\mathcal{A}^n$ were related to each other in such a way that all of them can be deduced from only one, say $\mathcal{C} \subseteq \mathcal{A}^n$, and if all of the decoding rules could be deduced from the decoding of $\mathcal{C}$.

This chapter concentrates on that method of construction. We assume that the set $\mathcal{A}$ is a group and then if $\mathcal{C}$ is a proper subgroup of $\mathcal{A}^n$, then the partition set (extracting function) is the quotient space $\mathcal{A}^n/\mathcal{C}$ and the embedding function maps any $(\mathbf{x}, \mathbf{s}) \in \mathcal{A}^n \times \mathcal{M}$ to an element in a coset of $\mathcal{A}^n/\mathcal{C}$ which is the inverse image of $\mathbf{s}$ by the extracting function. So the partitions that were central to the previous chapter appear in this chapter as the elements - cosets - of a quotient group.

3.1 Brief introduction to coding theory

Codes are used to correct errors introduced by transmission through a noisy channel. However, steganographic embedding schemes can be thought to introduce error in order to communicate a secret. The problem in coding theory is to find a code that can correct as much error as possible. However, the steganographic problem is to find a scheme that introduces small error. So we have interestingly divergent goals.

3.1.1 Basic definitions

We now proceed to the basic notions in coding theory.

Definition 3.1. *Let $\mathcal{A} = \{a_1, \ldots, a_q\}$ be an alphabet, whose elements are called* symbols. *A* block code *(or simply* code*) $\mathcal{C}$ of length n over $\mathcal{A}$ is a subset of $\mathcal{A}^n$. A sequence $\mathbf{c} \in \mathcal{C}$ is called a* codeword. *The number of elements of $\mathcal{C}$ is called the* size *of $\mathcal{C}$.*

A code of length n and size K is called an (n, K)-code.

Codes for which $\mathcal{A} = \mathbb{B}$ are called *binary codes*. In general if $|\mathcal{A}| = q$, then we refer to a q-ary code.

Definition 3.2. *If $\mathcal{A}$ is a group under the group operation $\star$, then a group code, $\mathcal{C}$, is a subgroup of the direct product group $\mathcal{A}^n$ under the componentwise group operation $\star$ such that for all $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} = (y_1, \ldots, y_n) : \mathcal{A}^n$,*

$$\mathbf{x} \star \mathbf{y} := (x_1 \star y_1, \ldots, x_n \star y_n).$$

Example 3.3. *Let $\mathcal{A} = \mathbb{B}$ and $n = 5$. Then the code $\mathbf{c} = \{\mathbf{c_0}, \mathbf{c_1}, \mathbf{c_2}, \mathbf{c_3}\} \subset \mathbb{B}^5$ such that*

$$\mathbf{c_0} = (0,0,0,0,0),\ \mathbf{c_1} = (1,1,1,0,0),\ \mathbf{c_2} = (0,1,1,1,1),\ \mathbf{c_3} = (1,0,0,1,1)$$

is a $(5,4)$-binary code. Moreover, Table 3.1 shows that it is a group code with the XOR operation on $\mathbb{B}$.

Table 3.1: Table of XOR on $\mathcal{C}$

$\oplus$	$(0,0,0,0,0)$	$(1,1,1,0,0)$	$(0,1,1,1,1)$	$(1,0,0,1,1)$
$(0,0,0,0,0)$	$(0,0,0,0,0)$	$(1,1,1,0,0)$	$(0,1,1,1,1)$	$(1,0,0,1,1)$
$(1,1,1,0,0)$	$(1,1,1,0,0)$	$(0,0,0,0,0)$	$(1,0,0,1,1)$	$(0,1,1,1,1)$
$(0,1,1,1,1)$	$(0,1,1,1,1)$	$(1,0,0,1,1)$	$(0,0,0,0,0)$	$(1,1,1,0,0)$
$(1,0,0,1,1)$	$(1,0,0,1,1)$	$(0,1,1,1,1)$	$(1,1,1,0,0)$	$(0,0,0,0,0)$

The rate of a code measure its transmission capacity, hence its efficiency.

Definition 3.4. *Let $\mathcal{C}$ be a q-ary (n, K)-code. Then the rate of $\mathcal{C}$ is defined by*

$$Rate(\mathcal{C}) := \frac{\log_q K}{n}. \tag{3.1.1}$$

Example 3.5. *The binary code $\mathcal{C}$ in Example 3.1 has rate*

$$\frac{\log_2 4}{5} = \frac{2}{5}.$$

The number of errors incurred in transmission is given by the Hamming distance between the transmitted and received words. The errors are the changes caused by the embedding process. Hamming distance is defined between two sequences of the same length to be the number of differences between them. It actually sums the per-symbol distance between the two sequences.

Definition 3.6. *Let $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$. Then for every $i \in [1, n]$ define*

$$\delta(x_i, y_i) := (x_i \neq y_i), \tag{3.1.2}$$

and define

$$d_H(\mathbf{x}, \mathbf{y}) := \sum_{i=1}^{n} \delta(x_i, y_i). \tag{3.1.3}$$

Definition 3.7. *The Hamming weight $w_H(\mathbf{x})$, of a word $\mathbf{x} = (x_1, \ldots, x_n)$ is the number of its non zero coordinates and it is defined as*

$$w_H(\mathbf{x}) := d_H(\mathbf{x}, \mathbf{0}) = \sum_{i=1}^{n} \delta(x_i, 0). \tag{3.1.4}$$

For a fixed length n, the Hamming distance is a metric on the vector space of words of that length.

Proposition 3.8. *For every $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{A}^n$, d_H satisfies the following*

1. $0 \leq d_H(\mathbf{x}, \mathbf{y}) \leq n$ *(non-negative and bounded)*
2. $d_H(\mathbf{x}, \mathbf{y}) = 0$ *if and only if* $\mathbf{x} = \mathbf{y}$ *(identity of indiscernibles)*
3. $d_H(\mathbf{x}, \mathbf{y}) = d_H(\mathbf{y}, \mathbf{x})$ *(symmetry)*
4. $d_H(\mathbf{x}, \mathbf{z}) \leq d_H(\mathbf{x}, \mathbf{y}) + d_H(\mathbf{y}, \mathbf{z})$ *(triangular inequality)*

Proof.

1. By definition, for all $i \in [1, n]$, $\delta(x_i, y_i) \in \{0, 1\}$. Hence
$$0 \leq \sum_{i=1}^{n} \delta(x_i, y_i) \leq n.$$

2. Two sequences of the same length are equal if and only if all coordinates are the same:
$$\begin{aligned} d_H(\mathbf{x}, \mathbf{y}) = 0 &\iff \forall i \in [1, n], \delta(x_i, y_i) = 0 \\ &\iff \mathbf{x} = \mathbf{y}. \end{aligned}$$

3. By definition, for all $i \in [1, n]$, $\delta(x_i, y_i) = \delta(y_i, x_i)$. Extending to d_H, we have
$$d_H(\mathbf{x}, \mathbf{y}) = d_H(\mathbf{y}, \mathbf{x}).$$

4. If we set $V(\mathbf{x}, \mathbf{z}) := \{i \in [1, n] \,|\, x_i \neq z_i\}$, then $d_H(\mathbf{x}, \mathbf{z}) = |V(\mathbf{x}, \mathbf{z})|$. For $\mathbf{y} \in \mathcal{A}^n$,
$$V(\mathbf{x}, \mathbf{z}) \subseteq V(\mathbf{x}, \mathbf{y}) \cup V(\mathbf{y}, \mathbf{z}).$$
Therefore we have
$$d_H(\mathbf{x}, \mathbf{z}) = |V(\mathbf{x}, \mathbf{z})| \leq |V(\mathbf{x}, \mathbf{y})| + |V(\mathbf{y}, \mathbf{z})| = d_H(\mathbf{x}, \mathbf{y}) + d_H(\mathbf{y}, \mathbf{z}).$$

□

The following is an interesting property of Hamming distance on commutative groups.

Lemma 3.9. *Let $(A, \star)$ be a finite Abelian group. Then the Hamming distance in $\mathcal{A}^n$ is translation invariant under $\star$, that is for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{A}^n$*
$$d_H(\mathbf{x}, \mathbf{y}) = d_H(\mathbf{x} \star \mathbf{z}, \mathbf{y} \star \mathbf{z}). \tag{3.1.5}$$
Moreover, for any subset $\mathcal{C} \subseteq \mathcal{A}^n$,
$$^1 d_H(\mathbf{x}, \mathbf{z} \star \mathcal{C}) = d_H(\mathbf{x} \star \mathbf{z}^{-1}, \mathcal{C}).$$

Proof. For $\mathbf{x}, \mathbf{y}, \mathbf{z} : \mathcal{A}^n$, we have
$$d_H(\mathbf{x}, \mathbf{y}) := |\{i \,|\, x_i \neq y_i\}|$$
and
$$d_H(\mathbf{x} \star \mathbf{z}, \mathbf{y} \star \mathbf{z}) := |\{i \,|\, x_i \star z_i \neq y_i \star z_i\}|.$$

[1] The Hamming distance from a point $\mathbf{x} \in \mathcal{A}^n$ to a subset $\mathcal{Y} \subseteq \mathcal{A}^n$ is
$$d_H(\mathbf{x}, \mathcal{Y}) = \min_{\mathbf{y} \in \mathcal{Y}} d_H(\mathbf{x}, \mathbf{y})$$
.

The equality $d_H(\mathbf{x},\mathbf{y}) = d_H(\mathbf{x}\star\mathbf{z},\mathbf{y}\star\mathbf{z})$ holds since $x_i \neq y_i \equiv x_i \star z_i \neq y_i \star z_i$ for all $i \in [1,n]$.

If $\mathcal{C} \subseteq \mathcal{A}^n$, then

$$\begin{aligned} d_H(\mathbf{x},\mathbf{z}\star\mathcal{C}) &:= \min_{\mathbf{c}\in\mathcal{C}} d_H(\mathbf{x},\mathbf{z}\star\mathbf{c}) && \text{by definition} \\ &= \min_{\mathbf{c}\in\mathcal{C}} d_H(\mathbf{x}\star\mathbf{z}^{-1},\mathbf{z}\star\mathbf{c}\star\mathbf{z}^{-1}) && \text{by group laws} \\ &= \min_{\mathbf{c}\in\mathcal{C}} d_H(\mathbf{x}\star\mathbf{z}^{-1},\mathbf{c}) && \text{by commutativity} \\ &= d_H(\mathbf{x}\star\mathbf{z}^{-1},\mathbf{c}). \end{aligned}$$

□

In terms of error correction, a good code can correct more error if the codewords are further apart (see Theorem 3.13). That means the distance between any two codewords must be as great as some constant $d \in [1,n]$, and that constant is called the distance of the code.

Definition 3.10. *The minimum distance (or distance) of a code $\mathcal{C}$ is the minimum between any two codewords of $\mathcal{C}$:*

$$d_H(\mathcal{C}) := \min\{d_H(\mathbf{c_1},\mathbf{c_2}) \mid \mathbf{c_1},\mathbf{c_2} \in \mathcal{C}, \mathbf{c_1} \neq \mathbf{c_2}\}. \tag{3.1.6}$$

A $(n,K)-$code with minimum distance d is called a (n,K,d)-code.

Example 3.11. *The minimum distance of the $(5,4)$-code $\mathcal{C}$ in given in Example 3.1 is $d_H(\mathcal{C}) = 3$ as shown in Table 3.2. Hence $\mathcal{C}$ is a $(5,4,3)$-code.*

Table 3.2: Distance between codewords of $\mathcal{C}$

d_H	$(0,0,0,0,0)$	$(1,1,1,0,0)$	$(0,1,1,1,1)$	$(1,0,0,1,1)$
$(0,0,0,0,0)$	-	3	4	3
$(1,1,1,0,0)$	3	-	3	4
$(0,1,1,1,1)$	4	3	-	3
$(1,0,0,1,1)$	3	4	3	-

We now show a connection between the distance of a code and the possibility of detecting and correcting errors.

Definition 3.12. *Let $\mathcal{C}$ be a code of length n over the alphabet $\mathcal{A}$.*

1. *$\mathcal{C}$ an r-*error detector[2] *if for every codeword $\mathbf{c} \in \mathcal{C}$ and every $\mathbf{x} \in \mathcal{A}^n$ with $\mathbf{x} \neq \mathbf{c}$, if $d_H(\mathbf{x},\mathbf{c}) \leq r$ then $\mathbf{x} \notin \mathcal{C}$.*

[2] r-error detector means can detect r errors.

2. $\mathcal{C}$ *a* t-error corrector[3] *if for every* $\mathbf{x} \in \mathcal{A}^n$, *if there exists* $\mathbf{c} \in \mathcal{C}$ *such that* $d_H(\mathbf{x}, \mathbf{c}) \leq t$ *then* $\mathbf{c}$ *is the unique closest codeword to* $\mathbf{x}$, *i.e.*

$$d_H(\mathbf{x}, \mathbf{c}) = d_H(\mathbf{x}, \mathcal{C})$$

and for any codeword $\mathbf{c}' \neq \mathbf{c}$,

$$d_H(\mathbf{x}, \mathbf{c}') > d_H(\mathbf{x}, \mathcal{C}).$$

The following theorem recasts that definition in terms of the minimum distance of the code.

Theorem 3.13. *Let* $\mathcal{C}$ *be a code of length* n *over the alphabet* $\mathcal{A}$.

1. $\mathcal{C}$ *is an* r*-error detector if and only if* $d_H(\mathcal{C}) > r$.

2. $\mathcal{C}$ *is a* t*-error corrector if and only if* $d_H(\mathcal{C}) \geq 2t + 1$.

Proof. 1. Let $\mathcal{C} \subseteq \mathcal{A}^n$ be a code that can detect r errors. Then by Definition 3.12, for all $\mathbf{c} \in \mathcal{C}$ and $\mathbf{x} \in \mathcal{A}^n$, $\mathbf{x} \neq \mathbf{c}$

$$\mathbf{x} \in \mathcal{C} \implies d_H(\mathbf{x}, \mathcal{C}) > r.$$

Conversely, if $d_H(\mathcal{C}) > r$, then for all $\mathbf{c}, \mathbf{c}' \in \mathcal{C}, \mathbf{c} \neq \mathbf{c}'$, $d_H(\mathbf{c}, \mathbf{c}') > r$. Thus if $d_H(\mathbf{c}, \mathbf{x}) \leq r$, then $\mathbf{x} \notin \mathcal{C}$.

2. Suppose that $\mathcal{C}$ can correct up to t errors. Then by Definition 3.12

$$d_H(\mathbf{x}, \mathbf{c}) \leq t \implies d_H(\mathbf{x}, \mathbf{c}) = d_H(\mathbf{x}, \mathcal{C}).$$

Conversely, if $d_H(\mathcal{C}) \geq 2t+1$, then spheres of radius t around codewords of $\mathcal{C}$ do not overlap. If $\mathbf{x}$ is in a sphere of radius $t \leq \frac{1}{2}(d-1)$ around a codeword $\mathbf{c}$ ($d_H(\mathbf{x}, \mathbf{c}) \leq t$), then $\mathbf{c}$ is the only codeword that satisfy

$$d_H(\mathbf{x}, \mathbf{c}) = d_H(\mathbf{x}, \mathcal{C}).$$

If $\mathbf{c}'$ is another codeword such that

$$d_H(\mathbf{x}, \mathbf{c}') = d_H(\mathbf{x}, \mathcal{C}),$$

then $d_H(\mathbf{x}, \mathbf{c}') > t$. Otherwise by the triangular inequality and $d_H(\mathbf{c}, \mathbf{c}') \geq d_H(\mathcal{C})$,

$$d_H(\mathbf{c}, \mathbf{c}') \leq d_H(\mathcal{C}, \mathbf{x}) + d_H(\mathbf{x}, \mathbf{c}') \leq 2t \leq d_H(\mathcal{C}) - 1 \quad \text{(contradiction)}.$$

□

[3] r-error corrector means can correct up to r errors.

Example 3.14. *Continuing Example 3.1, Table 3.2 shows that* $d_H(\mathcal{C}) = 3$*, therefore* $\mathcal{C}$ *detects* 2 *errors but can only correct* 1*.*

The aim of coding theory is to construct a code with a rate as close to 1 as possible and with as large a distance as possible. In other words, a good code has small n, large K and large d. If a code $\mathcal{C} \subseteq \mathcal{A}^n$ can correct up to t errors, then $\mathcal{C}$ is called a *t-error correcting code*. If $|\mathcal{A}| = q$, then $\mathcal{C}$ is a q-ary t-error correcting code.

We now give an upper bound limiting the maximum possible size of any code. The bound reflects that if $\mathcal{C}$ is a t-error correcting code, then if we place spheres of radius t around every codeword, the spheres must not overlap.

Theorem 3.15. *(Sphere-packing bound) (MacWilliams and Sloane, 1977) A t-error correcting code $\mathcal{C}$ of length n and cardinality K over an alphabet $\mathcal{A}$ with $q > 1$ elements must satisfy*

$$KV_q(t, n) \leq q^n. \tag{3.1.7}$$

Proof. Let C be a q-ary t-error correcting code of length n and cardinality K. Then any two spheres of radius t around distinct codewords are disjoint and there are K of them. Each of the K spheres contains $V_q(t, n)$ elements according to Lemma 3.20. All elements of $\mathcal{A}^n$ are not necessarily in the union of the K spheres. Therefore the cardinality of the union of the spheres around codewords of $\mathcal{C}$ must be less than or equal to q^n. □

There are codes that achieve the sphere-packing bound.

Definition 3.16. *An (n, K, t)-error correcting code $\mathcal{C}$ over an alphabet of size q is* perfect *if it satisfies the sphere-packing bound with equality.*

$$KV_q(t, n) = q^n. \tag{3.1.8}$$

Other parameters we can consider in this context are the average distance to code and the covering radius. Their definitions are given below.

Definition 3.17. *The average distance to a code $\mathcal{C}$ is denoted by $R_{\mathcal{C}}$ and given by*

$$R_{\mathcal{C}} := \frac{1}{q^n} \sum_{\mathbf{x} \in \mathcal{A}^n} d_H(\mathbf{x}, \mathcal{C}). \tag{3.1.9}$$

Definition 3.18. *The covering radius of a code $\mathcal{C} \subseteq \mathcal{A}^n$ is the smallest integer ρ such that the union of the spheres of radius ρ around the codewords of $\mathcal{C}$ cover the whole space $\mathcal{A}^n$. Thus*

$$\rho := min\{d \,|\, A^q \subseteq \cup_{\mathbf{c} \in \mathcal{C}} {}^4B(\mathbf{c}, d)\}. \tag{3.1.10}$$

[4] $B(\mathbf{c}, d)$ denotes the sphere of radius d around $\mathbf{c}$.

Thus every $\mathbf{x} \in \mathcal{A}^n$ *is at a distance at most* ρ *from a codeword of* $\mathcal{C}$*. That is*

$$\rho = \max_{\mathbf{x} \in \mathcal{A}^n} d_H(\mathbf{x}, \mathcal{C}) \leq n. \tag{3.1.11}$$

An (n, K)*-code with covering radius* ρ *is called a* (n, K, ρ)*-covering code*[5].

Definition 3.19. *Let* $\mathcal{A}$ *be an alphabet of size* q *with* $q > 1$*. Then for every* $\mathbf{x} \in \mathcal{A}^n$ *and every* $r \in \mathbb{N}$*, a sphere with center* $\mathbf{x}$ *and radius* r*, denoted* $B(\mathbf{x}, r)$*, is defined to be the set*

$$\{\mathbf{y} \in \mathcal{A}^n \mid d_H(\mathbf{x}, \mathbf{y}) \leq r\}.$$

The volume *of a sphere* $B(\mathbf{x}, r)$ *is denoted by* $V_q(n, r)$*.*

Lemma 3.20. *For every natural number* $r \geq 0$ *and alphabet* $\mathcal{A}$ *of size* $q > 1$*, and for every* $\mathbf{x} \in \mathcal{A}^n$*,* $B(\mathbf{x}, r)$ *contains*

$$V_q(n, r) = \sum_{i=0}^{r} \binom{n}{i} (q-1)^i. \tag{3.1.12}$$

elements if $r \leq n$*.*

Proof. Let $\mathbf{x} \in \mathcal{A}^n$. The number of vectors $\mathbf{y}$ at distance exactly i, $0 \leq i \leq n$, is equal to the number of ways to choose i positions in $\mathbf{x}$ to be changed and there are $q-1$ ways of changing each of these positions. Hence, the number of vectors at distance exactly i from $\mathbf{x}$ is $\binom{n}{i}(q-1)^i$. Since a sphere of radius r around $\mathbf{x}$ contains all vectors whose distance from $\mathbf{x}$ is in the range 0 to r, then the total number of vectors in a sphere of radius r around $\mathbf{x}$ is

$$\binom{n}{0} + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{r}(q-1)^R.$$

□

Observe that if $r > n$, then all vectors in the space are within distance r and so the sphere $B(\mathbf{x}, r)$ contains the entire space, hence $V_q(n, r) = q^n$. The following theorem gives a lower bound on the number of codewords of a code $\mathcal{C}$ with covering radius ρ (ρ-covering code), $0 \leq \rho \leq n$.

Theorem 3.21. *(Sphere-covering bound) A* (n, K, ρ)*-covering code over an alphabet* $\mathcal{A}$ *of size* $q > 1$ *satisfies*

$$K \geq \frac{q^n}{V_q(n, \rho)}. \tag{3.1.13}$$

[5] We use covering code when we want to specify that the third parameter is the covering radius.

Proof. Let $\mathcal{C} \subseteq \mathcal{A}^n$ be a (n, K, ρ)-covering code, $n, \rho \in \mathbb{N}$ and $0 \leq \rho \leq n$. Then for all $\mathbf{c} \in \mathcal{C}$, the sphere of radius ρ around $\mathbf{c}$ contains $V_q(n, \rho)$ elements, by Lemma 3.20. By Definition 3.18 of covering radius, we have

$$\mathcal{A}^n \subseteq \cup_{\mathbf{c}\in\mathcal{C}} B(\mathbf{c}, \rho).$$

Thus

$$q^n \leq \sum_{\mathbf{c}\in\mathcal{C}} |B(\mathbf{c}, \rho)|.$$

Since $|\mathcal{C}| = K$ and for all $\mathbf{c} \in \mathcal{C}$, $|B(\mathbf{c}, \rho)| = V_q(n, \rho)$ by Lemma 3.20, then we have

$$q^n \leq KV_q(n, \rho).$$

□

The equality in Equation 3.20 holds when any two distinct spheres of radius ρ around distinct codewords are disjoint. Perfect codes achieve the sphere-covering bound.

Lemma 3.22. *A t-error correcting code $\mathcal{C}$ is perfect if and only if its covering radius is $\rho = t$. That is, $\mathcal{C}$ can correct errors up to its covering radius.*

Proof. If $\rho = t$ then both Equation (3.1.7) and Equation (3.1.13) hold. Therefore

$$KV_q(n, t) = q^n.$$

Conversely, if we assume that $\mathcal{C}$ is perfect, then t is the smallest integer such that the union of the spheres of radius t around the codewords of $\mathcal{C}$ cover the whole space $\mathcal{A}^n$, which is exactly the meaning of covering radius. Therefore $\rho = t$. □

Example 3.23. *The code $\mathcal{C}$ in Example 3.1 is a $(5, 4, 2)-$covering code with minimum distance 3 (or $(5, 4, 3)$-code) over $\mathbb{B}$ and it has average distance $R_{\mathcal{C}} = 9/8$. Every $\mathbf{x} \in \mathbb{B}^5$ belongs to at least one sphere of radius 2 around the codewords of $\mathcal{C}$. The spheres are not pairwise disjoint because for example*

$$(1, 0, 1, 0, 1) \in B(\mathbf{c_1}, 2) \cap B(\mathbf{c_3}, 2).$$

$\mathcal{C}$ is not perfect.

3.1.2 Decoding

Let us consider $\mathcal{A}$ as an Abelian group with identity element[6] 0, and the all zero vector, denoted by $\mathbf{0}$, the identity element for $\mathcal{A}^n$. For a code[7] $\mathcal{C}$, if $\mathbf{x} \in \mathcal{A}^n$ is received, then decoding $\mathbf{x}$ means finding a closest $\mathbf{c} \in \mathcal{C}$. It is

[6] The group operation is in general addition. That is why we use 0.

[7] Here codes refers to group codes, i.e. subgroup of $\mathcal{A}^n$ (see Definition 3.2).

possible that there is more than one closest codeword. So decoding is a relation $Dec : \mathcal{A}^n \leftrightarrow \mathcal{C}$ which relates every $\mathbf{x} \in \mathcal{A}^n$ to each of its closest codewords. We denote $Dec_{\mathcal{C}}(\mathbf{x})$ the set of all closest codewords to $\mathbf{x}$.

So we formally define decoding as a *minimum distance decoding.*

Definition 3.24. *Let $\mathcal{C}$ be a code of length n over an alphabet $\mathcal{A}$. The* minimum distance decoding *rule states that every $\mathbf{x} \in \mathcal{A}^n$ is decoded to $\mathbf{c_x} \in \mathcal{C}$ that is closest to $\mathbf{x}$:*

$$Dec_{\mathcal{C}}(\mathbf{x}) = \{\mathbf{c_x} \in \mathcal{C} \mid d_H(\mathbf{x}, \mathbf{c_x}) = d_H(\mathbf{x}, \mathcal{C})\}. \tag{3.1.14}$$

A brute force algorithm for minimum distance decoding is given in Algorithm 3.1.

Algorithm 3.1 Minimum distance decoding algorithm (MDD)

1. Read the received vector $\mathbf{x} \in \mathcal{A}^n$ and the code $\mathcal{C}$.
2. Compute $d_H(\mathbf{x}, \mathbf{c})$ for all $\mathbf{c} \in \mathcal{C}$.
3. MDD decode $\mathbf{x}$ to $\mathbf{c_x}$ that is a closet codeword.

Efficiency of MDD: The sphere packing bound says that there are at most[8] $\frac{q^n}{q^t}$ codewords in $\mathcal{C}$, so the worst case running time for Step 1 and Step 2 is of $\mathcal{O}(nq^{n-t})$. Therefore the worst case running time is of $\mathcal{O}(nq^{n-t})$.

Example 3.25. *If the vector $\mathbf{x} = (0, 1, 0, 1, 1)$ is received, $Dec_{\mathcal{C}}(\mathbf{x}) = \{(0, 1, 1, 1, 1)\}$ (See Example3.1) because $(0, 1, 0, 1, 1) \in B(\mathbf{c_2}, 1)$. But if the vector $\mathbf{x}' = (1, 0, 1, 0, 1)$ is received then we can't correct $\mathbf{x}$ since for any $\mathbf{c} \in \mathcal{C}, \mathbf{x} \notin B(\mathcal{C}, 1)$. N can't decide between $\mathbf{c_1}$ and $\mathbf{c_3}$ which one is the correct sent codeword because*

$$d_H(\mathbf{x}', \mathcal{C}) = 2 = d_H(\mathbf{x}', \mathbf{c_1}) = d_H(\mathbf{x}', \mathbf{c_3}).$$

Decoding means deciding from a received $\mathbf{x}$ which codeword $\mathbf{c}$ was transmitted. But one can never be certain about $\mathbf{c}$. So a strategy is to find the most likely codeword $\mathbf{c}$, given that $\mathbf{x}$ is received. This strategy is called the *maximum likelihood decoding.*

Definition 3.26. *Let $\mathcal{C}$ be a code of length n over an alphabet $\mathcal{A}$. The* maximum likelihood decoding *rule states that every $\mathbf{x} \in \mathcal{A}^n$ is decoded to $\mathbf{c_x} \in \mathcal{C}$ when*

$$\Pr[\mathbf{x} \text{ received} \mid \mathbf{c_x} \text{ was sent}] = \max_{\mathbf{c} \in \mathcal{C}} \Pr[\mathbf{x} \text{ received} \mid \mathbf{c} \text{ was sent}]. \tag{3.1.15}$$

[8] Assuming that $\mathcal{C}$ is t-error correcting code and t is already known.

We assume that a codeword $\mathbf{c} = (c_1, \ldots, c_n) \in \mathcal{C}$ is transmitted. The number of errors in a received word $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{A}^n$ is equivalent to the distance (Hamming) between $\mathbf{c}$ and $\mathbf{x}$, where we define error as follows.

Definition 3.27. *The error vector is a vector* $\mathbf{e} = (e_1, \ldots, e_n)$ *such that for all* $i \in [1, n]$

$$x_i = c_i \star e_i, \tag{3.1.16}$$

or in other words,

$$e_i = x_i \star c_i^{-1}. \tag{3.1.17}$$

If a codeword $\mathbf{c}$ was transmitted without error, i.e. $\mathbf{e} = \mathbf{0}$, then the received vector $\mathbf{x} = \mathbf{c} \star \mathbf{0} = \mathbf{c}$ is *correct*.

Proposition 3.28. *Let* $\mathcal{C}$ *be a code, and a codeword* $\mathbf{c} \in \mathcal{C}$ *was sent. Then the received vector* $\mathbf{x}$ *is correct if and only if* $d_H(\mathbf{x}, \mathbf{c}) = \mathbf{0}$.

Proof. The received vector $\mathbf{x}$ is correct if and only if each symbol is correct, i.e. for all $i \in [1, n]$,

$$\begin{aligned} e_i = 0 &\iff x_i \star c_i^{-1} = 0 \\ &\iff x_i = c_i. \end{aligned}$$

By Proposition 3.8, $d_H(\mathbf{x}, \mathcal{C}) = \mathbf{0}$. □

Maximum likelihood decoding (Definition 3.26) finds the most likely error vector $\mathbf{e}$, given that $\mathbf{x}$ is received. Then decode $\mathbf{x}$ as $\mathbf{c} = \mathbf{x} \star \mathbf{e}^{-1}$. If we assign to each symbol of $\mathbf{x}$ a probability of being correct or not: $1 - p$ is the probability of x_i being correct (i.e. $e_i = 0$), where in general $0 \leq p < \frac{1}{2}$ (MacWilliams and Sloane, 1977), then we can deduce the following.

Proposition 3.29. *If the received vector* $\mathbf{x} \in \mathcal{A}^n$ *is not a codeword of* $\mathcal{C}$*, then the error vector* $\mathbf{e}$ *is a non zero vector of minimum weight in* $\mathcal{A}^n$ *such that* $\mathbf{x} \star \mathbf{e}^{-1} \in \mathcal{C}$.

Proof. Assume that error occurs with probability p independently at each symbol. That is for each $i \in [1, n]$,

$$Pr[e_i \neq 0] = p.$$

If $\mathbf{u} \in \mathcal{A}^n$ is a vector of weight a, then by independence

$$Pr[\mathbf{e} = \mathbf{u}] = p^a (1 - p)^{n-a}.$$

Since $p < 1/2$, the function $f(a) = p^a (1-p)^{n-a}$ decreases with $a \in [1, n]$. That is, the most likely error vector is of minimum weight. □

Algorithm 3.2 Maximum Likelihood Decoding Algorithm (MLD)

1. Read the received vector $\mathbf{x} \in \mathcal{A}^n$ and the code $\mathcal{C}$.
2. Find all vectors $\mathbf{u}$ such that $\mathbf{x} \star \mathbf{u}^{-1} \in \mathcal{C}$ (Non empty since $\mathbf{x}$ there).
3. In the set of all such $\mathbf{u}$, find one with smallest weight, and denote it $\mathbf{e}$.
4. MLD decode $\mathbf{x}$ to $\mathbf{c} = \mathbf{x} \star \mathbf{e}^{-1}$.

The maximum likelihood decoding algorithm finds the minimum weight vector that satisfies $\mathbf{x} \star \mathbf{e}^{-1} \in \mathcal{C}$ and then decodes $\mathbf{x}$ as $\mathbf{c} = \mathbf{x} \star \mathbf{e}^{-1}$.

Efficiency of MLD: Step1 is similar to MDD. For Step 2 we can find all the $\mathbf{u}$'s by computing $\mathbf{x} \star \mathcal{C}$ and that taking[9] $\mathcal{O}(nq^{n-t})$. Time for computing the weight and finding the smallest doesn't exceed $\mathcal{O}(nq^{n-t})$. Therefore the worst case running time is of $\mathcal{O}(nq^{n-t})$, similar to MDD.

We can improve Algorithm 3.2 by using the following lemma.

Lemma 3.30. *For a (n, M, ρ)-covering code $\mathcal{C} \subseteq \mathcal{A}^n$, the Hamming weight of an error vector is at most R.*

Proof. The sphere-covering bound (3.21) tells us that for every received vector $\mathbf{x} \in \mathcal{A}^n$ there exists $\mathbf{c} \in \mathcal{C}$ such that $\mathbf{x} \in B(\mathbf{c}, \rho)$. Moreover

$$\begin{aligned} d_H(\mathbf{x}, \mathbf{c}) &= d_H(\mathbf{x} \star \mathbf{c}^{-1}, 0) \\ &= w_H(\mathbf{x} \star \mathbf{c}^{-1}) \\ &\leq \rho. \end{aligned}$$

□

We assume that the covering radius of $\mathcal{C}$ is known[10].

Algorithm 3.3 Maximum Likelihood Decoding Algorithm (MLDI) improved

1. Read the received vector $\mathbf{x} \in \mathcal{A}^n$ and the code $\mathcal{C}$.
2. Find all vectors $\mathbf{u}$ such that $w_H(\mathbf{u}) \leq \rho$ and $\mathbf{x} \star \mathbf{u}^{-1} \in \mathcal{C}$.
3. In the set of all such $\mathbf{u}$, find the one with smallest weight, and denote it $\mathbf{e}$.
4. MLD decode $\mathbf{x}$ to $\mathbf{c} = \mathbf{x} \star \mathbf{e}^{-1}$.

[9] Running time is for the worst case.

[10] There should be algorithms to compute covering radius but here we assume it is known.

Efficiency of MLDI: For Step 2 we compute $\mathbf{x} \star \mathcal{C}$ and look in the set $\mathbf{u}$ with weight less than ρ the smallest weight. So running time for Step 3 is smaller but it still takes $\mathcal{O}(nq^n)$ time to run MLDI.

Now the decoding procedure is clear for a code $\mathcal{C}$. The next step consists of looking for a decoding map for the cosets of $\mathcal{C}$ that are in $\mathcal{A}^n/\mathcal{C}$.

3.1.3 Quotient space and cosets

Definition 3.31. *For a group code $\mathcal{C} \subseteq \mathcal{A}^n$. For each $\mathbf{z} \in \mathcal{A}^n$, the set*

$$\mathbf{z} \star \mathcal{C} := \{\mathbf{z} \star \mathbf{c} \,|\, \mathbf{c} \in \mathcal{C}\}$$

is called a coset *of $\mathcal{C}$. We define* quotient space

$$\mathcal{A}^n/\mathcal{C} := \{\mathbf{z} \star \mathcal{C} \,|\, \mathbf{z} \in \mathcal{A}^n\}.$$

A vector of minimum Hamming weight in a coset is called its leader. *It is possible that there are more than one vector of minimum weight in a coset, but choose one of them at random and call it the coset leader.*

The following proposition is given without proof, since it is basic group theory.

Proposition 3.32. *For any code*[11] *$\mathcal{C} \subset \mathcal{A}^n$, the following hold.*

1. *Each coset of $\mathcal{C}$ has cardinality equal to $\mathcal{C}$.*
2. *$\mathbf{z}$ is in the coset $\mathbf{z} \star \mathcal{C}$ for any $\mathbf{z} \in \mathcal{A}^n$.*
3. *For $\mathbf{x}, \mathbf{y} \in \mathcal{A}^n$, $\mathbf{x} \star \mathcal{C} = \mathbf{y} \star \mathcal{C}$ if and only if $\mathbf{x} \star \mathbf{y}^{-1} \in \mathcal{C}$.*
4. *For any $\mathbf{z} \in \mathcal{A}^n$ we have $\mathbf{z} \star \mathcal{C} = \mathcal{C}$ if and only if $\mathbf{z} \in \mathcal{C}$.*
5. *Two cosets are either disjoint or coincide, i.e. if $\mathbf{x}, \mathbf{y} \in \mathcal{A}^n$, then either $\mathbf{x} \star \mathcal{C} = \mathbf{y} \star \mathcal{C}$ or $(\mathbf{x} \star \mathcal{C}) \cap (\mathbf{y} \star \mathcal{C}) = \emptyset$.*

Note that if $|\mathcal{A}| = q$, the quotient group contains $q^n/|C|$ distinct cosets. Moreover if $\mathcal{C}$ is a (n, K, ρ)-covering code, then $\mathcal{A}^n/\mathcal{C}$ is a $(q^n/K, \rho)$-covering of $\mathcal{A}^n$ (see Chapter 2).

As in Example 3.33, the last two cosets in Table 3.3 have two minimum weight vectors and the others have exactly one. We denote by $\Omega_{\mathcal{C}}$ the set of all coset leaders of $\mathcal{C}$.

Now back to decoding. Suppose a vector $\mathbf{x} \in \mathcal{A}^n$ is received. Then $\mathbf{x}$ must belong to a coset of $\mathcal{C}$, say $\mathbf{x} = \mathbf{z} \star \mathcal{C}$ ($\mathbf{z} \in \mathcal{A}^n, \mathbf{c} \in \mathcal{C}$). If the codeword $\mathbf{c}'$ was sent, then the actual error vector $\mathbf{e} = \mathbf{x} \star \mathbf{c}'^{-1} = \mathbf{z} \star \mathcal{C} \star \mathbf{c}'^{-1} = \mathbf{z} \star \mathbf{c}'' \in \mathbf{z} \star \mathcal{C}$

[11] $\mathcal{C}$ is normal because it is a subgroup of a commutative group.

since $\mathcal{C}$ is a group code. Therefore the possible error vectors are in the coset containing $\mathbf{x}$.

MacWilliams and Sloane (1977) give a method of decoding by building the *standard array* table which consists of: the first row consists of the code itself, with the zero codeword on the left and the other rows being the other cosets $\mathbf{z} \star \mathcal{C}$, $\mathbf{z} \in \Omega_{\mathcal{C}}$, arranged in the same order and with the coset leader on the left.

Example 3.33. *Let $\mathcal{C}$ be the code in Example 3.1. Then the standard array table of $\mathcal{C}$ is given in Table 3.3.*

Table 3.3: A standard array for $\mathcal{C}$

$\mathcal{C} \oplus (0,0,0,0,0)$:	$(\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0})$	$(1,1,1,0,0)$	$(0,1,1,1,1)$	$(1,0,0,1,1)$
$\mathcal{C} \oplus (0,0,0,0,1)$:	$(\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{1})$	$(1,1,1,0,1)$	$(0,1,1,1,0)$	$(1,0,0,1,0)$
$\mathcal{C} \oplus (0,0,0,1,0)$:	$(\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{1},\mathbf{0})$	$(1,1,1,1,0)$	$(0,1,1,0,1)$	$(1,0,0,0,1)$
$\mathcal{C} \oplus (0,0,1,0,0)$:	$(\mathbf{0},\mathbf{0},\mathbf{1},\mathbf{0},\mathbf{0})$	$(1,1,0,0,0)$	$(0,1,0,1,1)$	$(1,0,1,1,1)$
$\mathcal{C} \oplus (0,1,0,0,0)$:	$(\mathbf{0},\mathbf{1},\mathbf{0},\mathbf{0},\mathbf{0})$	$(1,0,1,0,0)$	$(0,0,1,1,1)$	$(1,1,0,1,1)$
$\mathcal{C} \oplus (1,0,0,0,0)$:	$(\mathbf{1},\mathbf{0},\mathbf{0},\mathbf{0},\mathbf{0})$	$(0,1,1,0,0)$	$(1,1,1,1,1)$	$(0,0,0,1,1)$
$\mathcal{C} \oplus (0,0,1,0,1)$:	$(\mathbf{0},\mathbf{0},\mathbf{1},\mathbf{0},\mathbf{1})$	$(1,1,0,0,1)$	$(\mathbf{0},\mathbf{1},\mathbf{0},\mathbf{1},\mathbf{0})$	$(1,0,1,1,0)$
$\mathcal{C} \oplus (0,0,1,1,0)$:	$(\mathbf{0},\mathbf{0},\mathbf{1},\mathbf{1},\mathbf{0})$	$(1,1,0,1,0)$	$(\mathbf{0},\mathbf{1},\mathbf{0},\mathbf{0},\mathbf{1})$	$(1,0,1,0,1)$

We assume that the standard array is already given. The decoding using the standard array is called *standard array decoding* and it is given in Algorithm 3.4.

Algorithm 3.4 Standard array decoding (SAD)

1. Read the received vector $\mathbf{x} \in \mathcal{A}^n$.
2. Find the row of $\mathbf{x}$ in the standard array table.
3. Choose the error vector $\mathbf{e}$ as the coset leader found at the extreme left of $\mathbf{x}$.
4. SAD decode $\mathbf{x}$ to $\mathbf{c} = \mathbf{x} \star \mathbf{e}^{-1}$.

Efficiency of SAD: The time to find the row of $\mathbf{x}$ dominates the running time of SAD, and at most if is of $\mathcal{O}(q^n)$.

Example 3.34. *If we receive the vector $\mathbf{x} = (1,1,1,1,0)$, then by looking at the SA table in Example 3.33, the decoder decides that the error vector is $\mathbf{e} = (0,0,0,1,0)$ and then decodes $\mathbf{x}$ to $Dec_{\mathcal{C}}(\mathbf{x}) = \mathbf{x} \oplus \mathbf{e} = (1,1,1,0,0) = \mathbf{c_1}$.*

If a t-error correcting code $\mathcal{C}$ is perfect, then spheres of radius t around codewords do not overlap and cover the whole space $\mathcal{A}^n$. Therefore $\mathcal{C}$ can correct errors up to the covering radius $\rho = t$. Therefore $Dec_{\mathcal{C}}$ is a single function, and it maps each $\mathbf{x} \in \mathcal{A}^n$ to $\mathbf{c_x}$ such that $\mathbf{x} \in B(\mathbf{c_x}, t)$. That means that the error vector $\mathbf{e} = \mathbf{x} \star \mathcal{C}_{\mathbf{x}}^{-1}$ is unique (there is no other $\mathbf{e}' = \mathbf{x} \star \mathcal{C}^{-1}$ such that $w_H(\mathbf{e}) = w_H(\mathbf{e}')$). Thus the coset leaders of $\mathcal{C}$ are the unique vector of minimum weight in its coset. And by Lemma 3.30, if $\mathbf{z} \in \Omega_{\mathcal{C}}$ then $w_H(\mathbf{z}) \leq t$.

3.2 Stego-schemes from codes

In general codes are defined over the field $\mathbb{F}_q$ (q is a prime power) which, with the addition modulo q, is an Abelian group. Then we are first going to find the decoding map of the cosets of $\mathcal{A}^n/\mathcal{C}$, where $\mathcal{A}$ is an Abelian group and $\mathcal{C}$ is a group code defined over $\mathcal{A}$.

Let $\mathcal{A}$ be an Abelian group, $\mathcal{C} \subseteq \mathcal{A}^n$ be a group code (see Section 3.1.3) and $Dec_{\mathcal{C}}$ be decoding relation for $\mathcal{C}$. Then the following proposition gives the rules of decoding for the cosets of $\mathcal{C}$.

Proposition 3.35. *(Munuera, 2012) Let $\mathcal{A}$ be an Abelian group, $\mathcal{C} \subseteq \mathcal{A}^n$ a group code and $\mathbf{z} \in \mathcal{A}^n$. If $Dec_{\mathcal{C}}$ is a decoding for $\mathcal{C}$, then a decoding for the coset $\mathbf{z} \star \mathcal{C}$ relates any $\mathbf{x} \in \mathcal{A}^n$ to*

$$Dec_{\mathbf{z}\star\mathcal{C}}(\mathbf{x}) = \{\mathbf{z} \star \mathbf{c} \,|\, \mathbf{c} = Dec_{\mathcal{C}}(\mathbf{z}^{-1} \star \mathbf{x})\}.$$

Proof. It is well defined since $\mathbf{z} \star Dec_{\mathcal{C}}(\mathbf{x} \star \mathbf{z}^{-1}) \in \mathbf{z} \star \mathcal{C}$.

Now assume that there exists $\mathbf{y} \in \mathbf{z} \star \mathcal{C}$ such that

$$d_H(\mathbf{x}, \mathbf{y}) < d_H(\mathbf{x}, \mathbf{z} \star Dec_{\mathcal{C}}(\mathbf{z}^{-1} \star \mathbf{x})).$$

By Lemma 3.9, we have

$$d_H(\mathbf{z}^{-1} \star \mathbf{x}, \mathbf{z}^{-1} \star \mathbf{y}) < d_H(\mathbf{z}^{-1} \star \mathbf{x}, Dec_{\mathcal{C}}(\mathbf{z}^{-1} \star \mathbf{x})). \tag{3.2.1}$$

Since $\mathbf{z}^{-1} \star \mathbf{y} \in \mathcal{C}$, therefore the Inequality (3.2.1) contradicts the the definition $Dec_{\mathcal{C}}$, as minimum distance decoding. □

In order to define the embedding scheme, we need to describe the partition set index by $\mathcal{M}$. That is we need a group code $\mathcal{C} \subseteq \mathcal{A}^n$ of cardinality $\frac{q^n}{|\mathcal{M}|}$. So $|\mathcal{A}^n/\mathcal{C}| = |\mathcal{M}|$. Then we can define a one to one mapping $\phi : \mathcal{A}^n/\mathcal{C} \to \mathcal{M}$. Let $\Omega_{\mathcal{C}} = \{\mathbf{0}, \mathbf{z_1}, \mathbf{z_2}, \ldots, \mathbf{z}_{|\mathcal{M}|}\}$ be the set of coset leaders of $\mathcal{C}$. Then we have

$$\mathcal{A}^n/\mathcal{C} = \{\mathbf{z_1} \star \mathcal{C}, \mathbf{z_2} \star \mathcal{C}, \ldots, \mathbf{z}_{|\mathcal{M}|} \star \mathcal{C}\} \tag{3.2.2}$$

$$= \{[\mathbf{z_1}], [\mathbf{z_2}], \ldots, [\mathbf{z}_{|\mathcal{M}|}]\}. \tag{3.2.3}$$

Then equivalently $\phi : \Omega_{\mathcal{C}} \to \mathcal{M}$ where $\Omega_{\mathcal{C}}$ is the set of all coset leaders of $\mathcal{C}$ and each coset can be represented by its coset leader. If $\pi : \mathcal{A}^n \to \mathcal{A}^n/\mathcal{C}$

is the canonical projection map, then the extracting function is $Ext = \pi \circ \phi$. Therefore for every $\mathbf{s} \in \mathcal{M}$, we have $\mathcal{C}_{\mathbf{s}} = \phi^{-1}(\mathbf{s}) \star \mathcal{C}$. The embedding $\mathbf{s}$ maps any cover $\mathbf{x}$ to an element of $Dec_{\mathcal{C}_{\mathbf{s}}}$. The scheme $(Emb, Ext; \mathcal{A}^n, \mathcal{M})$ is a called *code-based stego-scheme* and it is proper.

Algorithm 3.5 describes the embedding and extracting from a code $\mathcal{C}$ which is a group code of length n on $\mathcal{A}$. The functions ϕ, π are as described above, the cover-sequence is from $\mathcal{A}^n$ and the secret to embedded is from $\mathcal{M}$.

Algorithm 3.5 Embedding scheme from a code $\mathcal{C}$ (CBE).

1. Given the code $\mathcal{C}$, the cover $\mathbf{x} \in \mathcal{A}^n$ and the secret $\mathbf{s} \in \mathcal{M}$.
2. Embedding: Modify the cover so that $\mathbf{y} = \phi^{-1}(\mathbf{s}) \star Dec_{\mathcal{C}}((\phi^{-1}(\mathbf{s}))^{-1} \star \mathbf{x})$.
3. Extraction: The secret is extracted such that $\mathbf{s} = \pi \circ \phi(\mathbf{y})$.

Efficiency of CBE: Embedding depends entirely on the decoding algorithm we choose for the code $\mathcal{C}$. So they takes equivalently the same amount of time.

To define a stego-scheme based on coding theory, we need a code and an efficient decoding algorithm on the code, such that its complexity is at most polynomial time.

3.3 Bounds on the parameters of code based stego-scheme

Some bounds on the parameters of code-steganographic schemes are given in this section. Those bounds are mostly derived from coding theoretic bounds.

The following proposition is the analogue of the Hamming bound in steganography. It says that the set of possible secrets that can be embedded is at most as great as the volume $V_q(n, R)$ of the sphere of radius R in $\mathcal{A}^q$ such that R is the embedding radius of the stego-scheme.

Proposition 3.36. *A (n, M, R)-embedding scheme*[12] *$\mathcal{S} = (Emb, Ext; \mathcal{X}, \mathcal{M})$ on $\mathcal{A}$ satisfies*

$$M \leq V_q(n, R). \tag{3.3.1}$$

Proof. Let $\mathbf{x} \in \mathcal{A}^n$. For any $\mathbf{s} \in \mathcal{M}$, we have $Emb(\mathbf{x}, \mathbf{s}) \in B(\mathbf{x}, R)$ by definition of embedding radius. By Proposition 2.5 in Chapter 2, for fixed $\mathbf{x} \in \mathcal{A}^n$, the map $Emb(\mathbf{x}, .) : \mathcal{M} \to B(\mathbf{x}, R)$ is injective. Therefore $|\mathcal{M}| \leq V_q(n, R)$. □

[12] R is here the embedding radius, and $M = |\mathcal{M}|$.

Schemes that achieve the bound (3.36) are called *maximum length embeddable code* (Zhang and Li, 2005) or *perfect* (Munuera, 2012).

Theorem 3.37. *If $\mathcal{C}$ is a (n, K, t)-error correcting*[13] *$(0 \leq t \leq n)$ perfect code over $\mathcal{A}$, then the code-based scheme $\mathcal{S}$ arising from $\mathcal{C}$ is a $(n, \frac{q^n}{K}, t)$ perfect stego-scheme.*

Proof. Let $\mathcal{C}$ be a t-error correcting perfect code of length n containing K codewords. Then the covering radius of $\mathcal{C}$ is $\rho = t$, which is the embedding radius of $\mathcal{C}$. That is $R = \rho = t$. Then $\mathcal{C}$ achieve the sphere covering bound (3.1.13), i.e.

$$q^n = KV_q(n, t) = KV_q(n, \rho) = KV_q(n, R)$$

Therefore $\mathcal{S}$ is perfect. The other parameters follows easily from the construction in Section 3.2. □

The relative payload or capacity is another important parameter of stego-schemes and it is defined as follows.

Definition 3.38. *The relative payload of a stego-scheme, denoted by α, is the number of embedded bits conveyed per single cover symbol. It is given by the ratio of the embedding capacity to the cover length. That is*

$$\alpha := \frac{h}{n}, \tag{3.3.2}$$

where $h = \log_2 |\mathcal{M}|$ is the embedding capacity defined in Chapter 2.

There is an obvious upper bound we can derive on the relative payload.

Proposition 3.39. *For a stego-scheme $(Emb, Ext; \mathcal{A}^n, \mathcal{M})$ such that $|\mathcal{A}| = q$,*

$$\alpha \leq \log_2 q. \tag{3.3.3}$$

Proof. Since the extracting function $Ext : \mathcal{A}^n \to \mathcal{M}$ is surjective by definition, then we have $|\mathcal{M}| \leq q^n$. Thus $\frac{\log_2 |\mathcal{M}|}{n} \leq \log_2 q$. □

A good scheme should have large α and e.

Lemma 3.40. *Let $q > 1$, n, be two integers and $0 < \rho \leq n - n/q$. Then*

$$V_q(n, \rho) \leq q^{nH_q\left(\frac{\rho}{n}\right)}, \tag{3.3.4}$$

where H_q is the q-ary entropy function defined by

$$H_q(p) = p \log_q(q-1) - p \log_q(p) - (1-p) \log_q(1-p). \tag{3.3.5}$$

[13] t-error correcting perfect codes are t-covering code (see Lemma 3.22).

Proof. By definition if the q-ary entropy, we have

$$H_q\left(\frac{\rho}{n}\right) = \frac{\rho}{n}\log_q(q-1) - \frac{\rho}{n}\log_q\left(\frac{\rho}{n}\right) - (1-\frac{\rho}{n})\log_q(1-\frac{\rho}{n}).$$

Therefore

$$\begin{aligned}
\frac{V_q(\rho,n)}{q^{nH_q\left(\frac{\rho}{n}\right)}} &= \frac{\sum_{i=0}^{\rho}\binom{n}{i}(q-1)^i}{(q-1)^\rho\left(1-\frac{\rho}{n}\right)^{\rho-n}\left(\frac{\rho}{n}\right)^{-\rho}} \\
&= \sum_{i=0}^{\rho}\binom{n}{i}(q-1)^i(q-1)^{-\rho}\left(1-\frac{\rho}{n}\right)^{n-\rho}\left(\frac{\rho}{n}\right)^{\rho} \\
&= \sum_{i=0}^{\rho}\binom{n}{i}(q-1)^i\left(1-\frac{\rho}{n}\right)^{n}\left(\frac{\frac{\rho}{n}}{(q-1)(1-\frac{\rho}{n})}\right)^{\rho} \\
&\leq \sum_{i=0}^{\rho}\binom{n}{i}(q-1)^i\left(1-\frac{\rho}{n}\right)^{n}\left(\frac{\frac{\rho}{n}}{(q-1)(1-\frac{\rho}{n})}\right)^{i} \\
&\leq \sum_{i=0}^{n}\binom{n}{i}\left(1-\frac{\rho}{n}\right)^{n-i}\left(\frac{\rho}{n}\right)^{i} \\
&= 1.
\end{aligned}$$

The first inequality comes from the fact that $\rho \leq n - n/q$ and hence

$$\frac{\frac{\rho}{n}}{(q-1)(1-\frac{\rho}{n})} \leq 1.$$

The second inequality is because $\rho \leq n$. □

From that lemma we can derive an upper bound on the embedding capacity.

Theorem 3.41. *The embedding capacity h of a q-ary[14] code-based stego-scheme $\mathcal{S} = (Emb, Ext; \mathcal{A}^n, \mathcal{M})$, with embedding radius[15] ρ satisfies the following inequality*

$$h \leq nH_q\left(\frac{\rho}{n}\right)\log_2 q. \tag{3.3.6}$$

Proof. The proof of this theorem follows easily from the previous Lemma 3.40:

$$\begin{aligned}
h &= \log_2|\mathcal{M}| && \text{by definition} \\
&\leq \log_2 V_q(n,\rho) && \text{by Prop. 3.36} \\
&= \log_q V_q(n,\rho)\log_2 q \\
&\leq nH_q\left(\frac{\rho}{n}\right)\log_2 q && \text{by Lemma 3.40.}
\end{aligned}$$

□

[14] q-ary stego-schemes refers to stego-schemes defined on $\mathcal{A}$ such that $|\mathcal{A}| = q$.

[15] We use $R = \rho$ for the embedding radius since our stego-scheme are based on a code of covering radius ρ. By Chapter 2, embedding radius and covering radius coincide.

Form Theorem 3.41, we can derive a bound on the relative capacity.

Corollary 3.42. *Let $\mathcal{S} = (Emb, Ext; \mathcal{A}^n, \mathcal{M})$ be a q-ary stego-scheme with embedding radius R. Then its relative payload satisfies*

$$\alpha \leq H_q\left(\frac{\rho}{n}\right)\log_2 q. \tag{3.3.7}$$

Proof. The proof follow easily from Theorem 3.41 and Definition 3.38 of relative payload. □

The following is an upper bound on the embedding efficiency.

Corollary 3.43. *(Fridrich* et al., *2007*a*) If $\mathcal{S} = (Emb, Ext; \mathcal{A}^n, \mathcal{M})$ be a q-ary stego-scheme with relative message length α, then the following upper bound holds for its lower embedding efficiency*

$$\underline{e} \leq \frac{\alpha}{H^{-1}\left(\frac{\alpha}{\log_2 q}\right)}. \tag{3.3.8}$$

Proof. We have $\underline{e} := \frac{h}{\rho}$ and $\alpha := \frac{h}{n}$. Thus

$$\underline{e} = \frac{\alpha n}{\rho}. \tag{3.3.9}$$

Moreover, from Corollary 3.42, we have

$$H_q^{-1}\left(\frac{\alpha}{\log_2 q}\right) \leq \frac{\rho}{n}, \tag{3.3.10}$$

where H_q^{-1} is the inverse function of the H_q (see Appendix **??**). Therefore, from Inequalities (3.3.9) and (3.3.10),

$$\underline{e} \leq \frac{\alpha}{H^{-1}\left(\frac{\alpha}{\log_2 q}\right)}. \tag{3.3.11}$$

□

We compare schemes having the same relative payload by their embedding efficiencies. So scheme that achieve (if possible) or at least as close as the bound (3.43) is preferable. Fridrich *et al.* (2007*a*) state that there exist stego-schemes based on linear codes whose lower embedding efficiency is asymptotically optimal, i.e. achieve the upper bound (3.43).

Chapter 4

Steganographic Scheme from linear codes: Matrix embedding

In this chapter we connect the previous two chapters. In particular we specialise the theory of Chapter 3 to "linear" codes to show that some bounds in Chapter 3 can be achieved by random linear codes. We also establish that the result is best possible. A linear code enables us to encrypt and decrypt by a linear transformation and hence by multiplying by a matrix.

A particular case of a code-based stego-scheme arising from linear codes is called *matrix embedding*. It was first introduced by Crandall (1998). It requires the sender and recipient to agree in advance on a parity check matrix $\mathbf{H}$, and then the secret is extracted by the recipient as the syndrome with respect to $\mathbf{H}$ of the received cover object. This method is popular because of the $F5$ algorithm of (Westfeld, 2001), which can embed t bits of message in $2^t - 1$ cover symbols by changing at most one of them. The $F5$ algorithm is a specific implementation of matrix encoding by using Hamming codes. That is why we can directly explain the parameters above, since Hamming codes are of length $2^t - 1$, redundancy t and covering radius 1.

4.1 Linear codes

This section recalls several notions on linear codes.

Definition 4.1. *A $[n, k]_q$ linear code $\mathcal{C}$ is a k-dimensional subspace of an n-dimensional vector space over a finite field $\mathbb{F}_q$ (q a prime power).*

A $[n, k]_q$ linear code $\mathcal{C}$ can be represented as the null space of a matrix[1] $\mathbf{H} \in (\mathbb{F}_q)_{(n-k)\times n}$. Such a matrix is called a *parity check matrix* of $\mathcal{C}$ (MacWilliams and Sloane, 1977).

There are several consequences of a code being linear. Let $\mathbf{H}$ be a parity check matrix of a linear code $\mathcal{C}$:

[1] We denote by $(\mathbb{F}_q)_{(n-k)\times n}$ the set of all $(n-k) \times n$ matrices on $\mathbb{F}_q$.

1. If $\mathbf{x}$ and $\mathbf{y}$ are codewords of $\mathcal{C}$, then so is[2] $\mathbf{x}+\mathbf{y}$, because

$$(\mathbf{x}+\mathbf{y})\mathbf{H}^{tr} = \mathbf{x}\mathbf{H}^{tr} + \mathbf{y}\mathbf{H}^{tr} = \mathbf{0}.$$

2. If $c \in \mathbb{F}_q$, then $c\mathbf{x}$ is also a codeword, because $(c\mathbf{x})\mathbf{H}^{tr} = c(\mathbf{x}\mathbf{H}^{tr}) = \mathbf{0}$.

Moreover, a linear code forms an additive group and a vector space over the field $\mathbb{F}_q$. Hence linear codes over $\mathbb{F}_q$ are group codes[3].

To find the minimum distance for a general code we have to look at all pairs of codewords. But for linear codes there's a short-cut.

Proposition 4.2. *The minimum distance of a linear code is equal to the minimum weight of its non zero codewords.*

Proof. Suppose two distinct codewords $\mathbf{c},\ \mathbf{c}' \in \mathcal{C}$ such that $d_H(\mathbf{c},\mathbf{c}') = d$. Then we have $w_H(\mathbf{c}-\mathbf{c}') = d$. $\mathbf{c}-\mathbf{c}'$ is a codeword of $\mathcal{C}$ because

$$(\mathbf{c}-\mathbf{c}')\mathbf{H}^{tr} = \mathbf{c}\mathbf{H}^{tr} - \mathbf{c}'\mathbf{H}^{tr} = \mathbf{0}$$

Conversely, if $\mathbf{c} \neq \mathbf{0}$ is a codeword such that $w_H(\mathbf{c}) = d$, then the minimum distance of $\mathcal{C}$ is at most d, since $\mathbf{0} \in \mathcal{C}$ and $d_H(\mathbf{c},\mathbf{0}) = w_H(\mathbf{c})$. □

An $[n,k]_q$ linear code $\mathcal{C}$ with minimum distance (or weight) d is called an $[n,k,d]_q$-code[4].

Example 4.3. *A parity check matrix for $\mathcal{C}$ (in Example 3.1) is given by*

$$\mathbf{H} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{pmatrix}.$$

It easy to see that $\mathcal{C}$ is the null space of $\mathbf{H}$. In addition, the code $\mathcal{C}$ is a binary linear $[5,2,3]$-code, where 5 is the length, 2 the dimension and 3 the distance.

4.2 Matrix embedding theorem

To define a matrix embedding scheme we need a linear code $\mathcal{C}$, and a decoding algorithm on $\mathcal{C}$. We follow the concepts of Chapter 3. As partition of $\mathbb{F}_q^n$ we consider the quotient space $\mathbb{F}_q^n/\mathcal{C}$ such that $\mathcal{C}$ is an $[n,k]$ linear code over $\mathbb{F}_q$. The elements of the quotient space are the cosets of $\mathcal{C}$ and are defined as follows for any $[n,k]_q$ linear code.

[2] $\mathbf{x}+\mathbf{y} = (x_i + y_i \,|\, i \in (0,n])$ if $\mathbf{x} = (x_i \,|\, i \in (0,n])$ and $\mathbf{y} = (y_i \,|\, i \in (0,n])$, where the addition is modulo q.

[3] Group codes are not necessarily linear.

[4] The (n,M,d) notation for general codes is generally replaced by $[n,k,d]$ for linear codes since k is here the dimension of the subspace. We write $[n,k,d]_q$ to specify that the code is over the field $\mathbb{F}_q$.

Definition 4.4. *Let $\mathcal{C}$ be an $[n,k]_q$ linear code. For any $\mathbf{z} \in \mathbb{F}_q^n$, the set*

$$\mathbf{z} + \mathcal{C} = \{\mathbf{z} + \mathbf{c} \,|\, \mathbf{c} \in \mathcal{C}\}$$

is called a coset *of $\mathcal{C}$.*

Here are some facts about cosets of a linear code, given without proof as an analogue of Proposition 3.32.

Proposition 4.5. *Let $\mathcal{C}$ be an $[n,k]_q$ linear code. Then*

1. *Every vector $\mathbf{x}$ must be in some coset of $\mathcal{C}$ (e.g. $\mathbf{x} \in \mathbf{x} + \mathcal{C}$).*
2. *Two vectors $\mathbf{x}$ and $\mathbf{y}$ are in the same coset if and only if $(\mathbf{x} - \mathbf{y}) \in \mathcal{C}$.*
3. *We have $\mathbf{z} + \mathcal{C} = \mathcal{C}$ if and only if $\mathbf{z} \in \mathcal{C}$.*
4. *Two cosets are either disjoint or coincide.*

We represent cosets by their minimum weight vectors called *leaders*. That is, we write $\mathbf{z} + \mathcal{C}$, for a coset with leader $\mathbf{z}$. So finding which coset a vector $\mathbf{y}$ is in, means finding the vector $\mathbf{z}$ such that $\mathbf{y} \in \mathbf{z} + \mathcal{C}$. There is an easy way to do that: by computing the *syndrome* of $\mathbf{y}$ with respect to a parity check matrix of $\mathcal{C}$.

Definition 4.6. *Let $\mathcal{C}$ be a linear $[n,k,d]$-code over $\mathbb{F}_q$ and let $\mathbf{H}$ be a parity-check matrix for $\mathcal{C}$. Then for every $\mathbf{x} \in \mathbb{F}_q^n$ the* syndrome *of $\mathbf{x}$ determined by $\mathbf{H}$, denoted by $\mathbf{s}_\mathbf{H}(\mathbf{x})$, is defined as*

$$\mathbf{s}_\mathbf{H}(\mathbf{x}) := \mathbf{x}\mathbf{H}^{tr} \in \mathbb{F}_q^{n-k}.$$

Proposition 4.7. *Let $\mathcal{C}$ be a linear $[n,k]$-code over $\mathbb{F}_q$ and let $\mathbf{H}$ be a parity-check matrix for $\mathcal{C}$. Then for every $\mathbf{x}, \mathbf{y} \in \mathbb{F}_q^n$:*

1. $\mathbf{s}_\mathbf{H}(\mathbf{x} + \mathbf{y}) = \mathbf{s}_\mathbf{H}(\mathbf{x}) + \mathbf{s}_\mathbf{H}(\mathbf{y})$,
2. $\mathbf{x} \in \mathcal{C}$ *if an only if* $\mathbf{s}_\mathbf{H}(\mathbf{x}) = \mathbf{0}$,
3. $\mathbf{s}_\mathbf{H}(\mathbf{x}) = \mathbf{s}_\mathbf{H}(\mathbf{y})$ *if and only if* $\mathbf{x}$ *and* $\mathbf{y}$ *are in the same coset.*

Proof.

1. $\mathbf{s}_\mathbf{H}(\mathbf{x} + \mathbf{y}) = \mathbf{x}\mathbf{H}^{tr} + \mathbf{y}\mathbf{H}^{tr} = \mathbf{s}_\mathbf{H}(\mathbf{x}) + \mathbf{s}_\mathbf{H}(\mathbf{y})$.
2. $\mathbf{x} \in \mathcal{C}$ if and only if $\mathbf{x}\mathbf{H}^{tr} = \mathbf{0}$ if and only if $\mathbf{s}_\mathbf{H}(\mathbf{x}) = \mathbf{0}$.
3. $\mathbf{s}_\mathbf{H}(\mathbf{x}) = \mathbf{s}_\mathbf{H}(\mathbf{y})$ if and only if $\mathbf{x}\mathbf{H}^{tr} = \mathbf{y}\mathbf{H}^{tr}$, i.e. $(\mathbf{x} - \mathbf{y})\mathbf{H}^{tr} = \mathbf{0}$. Which is true if and only if $\mathbf{x} - \mathbf{y} \in \mathcal{C}$. Equivalently, by Proposition 4.5 $\mathbf{x}$ and $\mathbf{y}$ are in the same coset.

□

As a consequence of Proposition 4.7, we can deduce the following theorem.

Theorem 4.8. *If $\mathcal{C}$ is a linear $[n,k]_q$-code, with parity check matrix $\mathbf{H}$, then for each $\mathbf{s} \in \mathbb{F}_q^{n-k}$, the set*

$$\mathcal{C}(\mathbf{s}) = \{\mathbf{x} \in \mathbb{F}_q^n \,|\, \mathbf{x}\mathbf{H}^{tr} = \mathbf{s}\}$$

coincides with a coset of $\mathcal{C}$ and we have

$$\mathbb{F}_q^n = \sqcup_{\mathbf{s}\in\mathbb{F}_q^{n-k}} \mathcal{C}(\mathbf{s}).$$

Proof. Let $\mathbf{z}+\mathcal{C}$ be a coset and $\mathbf{z}\mathbf{H}^{tr} = \mathbf{s}$. By Proposition 4.7: $\mathbf{x} \in \mathbf{z}+\mathcal{C}$ if and only if $\mathbf{x}\mathbf{H}^{tr} = \mathbf{z}\mathbf{H}^{tr} = \mathbf{s}$. Thus $\mathbf{x} \in \mathcal{C}(\mathbf{s})$. □

Let $\Omega_{\mathcal{C}}$ be the set of all coset leaders of $\mathcal{C}$.

Proposition 4.9. *If $\mathcal{C}$ is a linear code with covering radius ρ, then for every $\mathbf{z} \in \Omega_{\mathcal{C}}$,*

$$w_H(\mathbf{z}) \leq \rho.$$

Proof. Let $\mathbf{x} \in \mathbb{F}_q^n$ and $\mathbf{s} = \mathbf{x}\mathbf{H}^{tr}$. Then we have

$$\mathcal{C}(\mathbf{s}) = \{\mathbf{y} \in \mathbb{F}_q^n \,|\, \mathbf{y} = \mathbf{x} - \mathbf{c},\, \mathbf{c} \in \mathcal{C}\}.$$

If $\mathbf{z}$ is a coset leader, then By definition

$$\begin{aligned} \rho &:= \max_{\mathbf{u}\in\mathbb{F}_q^n} d_H(\mathbf{u}, \mathcal{C}) \\ &\geq d_H(\mathbf{z}, \mathcal{C}) \\ &= \min_{\mathbf{c}\in\mathcal{C}} w_H(\mathbf{x}-\mathbf{c}) | \mathbf{c} \in \mathcal{C}\} \\ &= w_H(\mathbf{z}). \end{aligned}$$

□

If $\mathcal{C}$ is a linear code with parity check matrix $\mathbf{H}$, a table containing all the pairs $(\mathbf{z}, \mathbf{s}_{\mathbf{H}}(\mathbf{z}))$, where $\mathbf{z}$ is a coset leader, is called a *syndrome decoding array* (SDA).

Example 4.10. *Table 4.1 gives the SDA table for the code $\mathcal{C}$ in Example 3.1.*

For a linear code $\mathcal{C}$, the SDA is built especially for decoding. A decoding algorithm for linear code $\mathcal{C}$ is given in Algorithm 4.2, and called the *Standard array decoding algotithm* (SAD).

Efficiency of SDA: Sorting the SDA, this procedure takes time $\mathcal{O}(nk)$. By traversing all the cosets and computing the leader and its syndrome, an SAD takes $\mathcal{O}(q^n)$ time.

If we denote by $cl(\mathbf{s})$ a leader of the coset $\mathcal{C}(\mathbf{s})$, then we can recapitulate SDA as:

$$Dec_{\mathcal{C}}(\mathbf{x}) = \mathbf{x} - cl(\mathbf{s}_{\mathbf{H}}(\mathbf{x}))$$

Table 4.1: The syndrome decoding array (SDA)

$\mathbf{z}$	$\mathbf{s} = \mathbf{z}\mathbf{H}^{tr}$
$(0,0,0,0,0)$	$(0,0,0)$
$(0,0,0,0,1)$	$(1,1,0)$
$(0,0,0,1,0)$	$(1,0,1)$
$(0,0,1,0,0)$	$(0,1,0)$
$(0,1,0,0,0)$	$(0,0,1)$
$(1,0,0,0,0)$	$(0,1,1)$
$(0,0,1,0,1)$	$(1,0,0)$
$(0,0,1,1,0)$	$(1,1,1)$

Algorithm 4.1 Standard array decoding algorithm (SAD)

1. Given a vector $\mathbf{x} \in \mathbb{F}_q^n$, compute $\mathbf{s}_\mathbf{H}(\mathbf{x})$.
2. Find the coset leader $\mathbf{z}$ of $\mathcal{C}(\mathbf{s}_\mathbf{H}(\mathbf{x}))$ by looking up $\mathbf{s}_\mathbf{H}(\mathbf{x})$ in the SDA.
3. Output: $\mathbf{x} - \mathbf{z}$.

for all $\mathbf{x} \in \mathbb{F}_q^n$, a code linear code $\mathcal{C}$ and a parity check matrix $\mathbf{H}$ for $\mathcal{C}$.

To define a stego-scheme from a linear code $\mathcal{C}$, it remains for us to define a decoding rule for the cosets of $\mathcal{C}$. By Proposition 3.35: if $\mathbf{z} + \mathcal{C}$ is a coset of $\mathcal{C}$, then a decoding for $\mathbf{z} + \mathcal{C}$ is

$$Dec_{\mathbf{z}+\mathcal{C}}(\mathbf{x}) = \mathbf{z} + Dec_{\mathcal{C}}(\mathbf{x} - \mathbf{z}).$$

Translating to embedding scheme we have, for all secrets $\mathbf{s} \in \mathbb{F}_q^{n-k}$ and covertext $\mathbf{x} \in \mathbb{F}_q^n$

$$\begin{aligned} Emb(\mathbf{x}, \mathbf{s}) &= cl(\mathbf{s}) + Dec_{\mathcal{C}}(\mathbf{x} - cl(\mathbf{s})) \\ &= cl(\mathbf{s}) + \mathbf{x} - cl(\mathbf{s}) - cl(\mathbf{s}_\mathbf{H}(\mathbf{x} - cl(\mathbf{s}))) \\ &= \mathbf{x} - cl(\mathbf{s}_\mathbf{H}(\mathbf{x}) - \mathbf{s}). \end{aligned}$$

Theorem 4.11. *(Fridrich* et al.*, 2007a) Let $\mathcal{C}$ be a linear $[n,k]_q$ linear code with parity check matrix $\mathbf{H}$ and covering radius ρ. The embedding scheme below can communicate $n-k$ symbols in a sequence $\mathbf{x}$ of length n ($\mathbf{x} \in \mathbb{F}_q^n$) using at most ρ changes:*

$$Emb(\mathbf{x}, \mathbf{s}) = \mathbf{x} - cl(\mathbf{s}_\mathbf{H}(\mathbf{x}) - \mathbf{s}) = \mathbf{y},$$

$$Ext(\mathbf{y}) = \mathbf{s}_\mathbf{H}(\mathbf{y})$$

for $\mathbf{s} \in \mathbb{F}_q^{n-k}$ and $\mathbf{x} \in \mathbb{F}_q^n$.

Proof. We first prove that the stego-scheme of Theorem 4.11 is well defined, i.e. satisfies $Ext(Emb(\mathbf{x}, \mathbf{s})) = \mathbf{s}$.

$$\begin{aligned} Ext(Emb(\mathbf{x}, \mathbf{s})) &= \mathbf{y}\mathbf{H}^{tr} \\ &= \mathbf{x}\mathbf{H}^{tr} - (cl(\mathbf{s}_{\mathbf{H}}(\mathbf{x}) - \mathbf{s}))\mathbf{H}^{tr} \\ &= \mathbf{x}\mathbf{H}^{tr} - \mathbf{x}\mathbf{H}^{tr} + \mathbf{s} \\ &= \mathbf{s}. \end{aligned}$$

Since the code $\mathcal{C}$ is a $[n, k]$-code with covering radius ρ, by Proposition 4.9 we can deduce that

$$d_H(\mathbf{x}, \mathbf{y}) = w_H(cl(\mathbf{s}_{\mathbf{H}}(\mathbf{x}) - \mathbf{s})) \leq \rho,$$

which proves that the changes are at most ρ. □

A steganographic scheme using matrix embedding from a linear $[n, k]_q$-code $\mathcal{C}$ with parity check matrix $\mathbf{H}$ is given by Algorithm 4.2.

Algorithm 4.2 Matrix embedding using linear codes (ME)

1. Given the cover-text $\mathbf{x} \in \mathbb{F}_q^n$, the secret $\mathbf{s} \in \mathbb{F}_q^{n-k}$ and a SDA for $\mathcal{C}$.
2. Compute the syndrome $\mathbf{s}_{\mathbf{H}}(\mathbf{x}) = \mathbf{x}\mathbf{H}^{tr}$.
3. Find in the SDA the leader $\mathbf{z}$ corresponding to the syndrome $\mathbf{s}_{\mathbf{H}}(\mathbf{x}) - \mathbf{s} \in \mathbb{F}_q^{n-k}$.
4. Embedding: Compute $\mathbf{y} = \mathbf{x} - \mathbf{z}$.
5. Extraction: Compute the syndrome $\mathbf{s}_{\mathbf{H}}(\mathbf{y}) = \mathbf{y}\mathbf{H}^{tr}$.

Efficiency of ME: Similarly to SAD, ME takes time of $\mathcal{O}(q^n)$.

Example 4.12. *Let us use the linear $[5, 2]_2$-code $\mathcal{C}$ (of Example 3.1), with the parity check matrix $\mathbf{H}$ given in Example 4.3, to embed 3 bits in a sequence in $\mathbb{B}^5$. If $\mathbf{s} = (0, 1, 1)$ is the secret to embedded in the cover-text $\mathbf{x} = (0, 0, 0, 1, 0)$ to embed $\mathbf{s}$, then the Algorithm 4.2 output*

$$\mathbf{y} = (0, 0, 0, 1, 0) - cl((1, 0, 1) - (0, 1, 1)) = (0, 0, 0, 1, 0) - cl(1, 1, 0) = (0, 0, 0, 1, 1).$$

And we can recover $\mathbf{s}$ such that

$$\mathbf{s}_{\mathbf{H}}(\mathbf{y}) = \mathbf{y}\mathbf{H}^{tr} = (0, 1, 1) = \mathbf{s}.$$

4.3 Parameters of Linear Stego-scheme

The performance of a steganographic method can be measured in terms of the embedding efficiency which is defined as follows.

Definition 4.13. *The* embedding efficiency *of stego-scheme is the expected number of embedded random messages bits per one embedding changes. It is denoted by e and defined by the ratio between the number of message symbols we can embed and the average number of embedding changes, i.e.*

$$e := \frac{k}{R_a} \tag{4.3.1}$$

where

$$R_a = \sum_{\mathbf{x},\mathbf{s}} d_H(\mathbf{x}, Emb(\mathbf{x}, \mathbf{s})). \tag{4.3.2}$$

The following theorem gives the relative payload and the embedding efficiency of a matrix embedding scheme form a linear code.

Theorem 4.14. *The relative payload and the embedding efficiency of a linear stego-scheme from a linear $[n,k]_q$-code $\mathcal{C}$ are respectively*

$$\alpha = \frac{n-k}{n}$$

and

$$e = \frac{n-k}{R_{\mathcal{C}}}$$

where $R_{\mathcal{C}}$ is the average distance to the code $\mathcal{C}$.

Proof. The proof of the relative payload follows easily from the definition of matrix embedding scheme. To prove the embedding efficiency we need to prove that the average number of embedding changes introduced by the embedding function is equal to the average distance to the code. The average number of embedding changes, for uniformly chosen secret $\mathbf{s} \in \mathbb{F}_q^{n-k}$, is

$$\begin{aligned}
\frac{1}{q^{n-k}} \sum_{\mathbf{s}\in\mathbb{F}_q^{n-k}} w_H(cl(\mathbf{s})) &= \frac{1}{q^n} \sum_{\mathbf{s}\in\mathbb{F}_q^{n-k}} q^k w_H(cl(\mathbf{s})) \\
&= \frac{1}{q^n} \sum_{\mathbf{s}\in\mathbb{F}_q^{n-k}} \sum_{\mathbf{x}\in\mathcal{C}(\mathbf{s})} w_H(cl(\mathbf{s})) \\
&= \frac{1}{q^n} \sum_{\mathbf{s}\in\mathbb{F}_q^{n-k}} \sum_{\mathbf{x}\in\mathcal{C}(\mathbf{s})} d_H(\mathbf{x}, \mathcal{C}) \\
&= \frac{1}{q^n} \sum_{\mathbf{x}\in\mathbb{F}_q^{n}} d_H(\mathbf{x}, \mathcal{C}) \\
&= R_{\mathcal{C}}.
\end{aligned}$$

□

The following gives the parameters of a stego-scheme from the binary Hamming code of length $n = 2^r - 1$ ($r \geq 2$), dimension $k = 2^r - 1 - r$ and covering radius 1. MacWilliams and Sloane (1977) give more details about Hamming codes.

Example 4.15. *(Fridrich and Soukal, 2006) Let* $\mathbf{H}$ *be the check matrix of the binary* $[2^r - 1, 2^r - 1 - r]$ *Hamming code* $\mathcal{H}_r$*. The parameters of the matrix embedding scheme form* $\mathcal{H}_k$ *verifies:*

$$\alpha = \frac{r}{2^r - 1}, \; e = \frac{r}{1 - 2^{-r}}.$$

Table 4.2: Relative and embedding efficiency for Hamming code-based steganography

k	Relative Payload	Embedding efficiency
1	1.000	2.000
2	0.667	2.667
3	0.429	3.429
4	0.267	4.267
5	0.161	5.161
6	0.093	6.093
7	0.055	7.055
8	0.031	8.031
9	0.018	9.018

We can see that embedding efficiency increases with r while relative payload decreases (see Table 4.2). Hamming codes are well suited when the size of the secret to be embedded is a small fraction of the cover-text, since many bits can be embedded with a single change (Hamming code are single error-correcting codes).

There are linear stego-scheme that are perfect. By Theorem 3.37 they arise from perfect linear codes. Tietäväinen (1973) shows that there are only three kinds of trivial perfect codes:

1. The q-ary $\left[\frac{q^r-1}{q-1}, \frac{q^r-1}{q-1} - r\right]$ Hamming codes.
2. The binary $[23, 12]$ Golay code.
3. The ternary $[11, 6]$ Golay code.

From those codes are built the perfect linear stego-schemes (Zhang and Li, 2005)):

1. The binary linear $(23, 2^{11}, 3)$ code.
2. The ternary linear $(11, 3^5, 2)$ code.
3. The $\left(\frac{q^r-1}{q-1}, q^r, 1\right)$ code over $\mathbb{F}_q$.

And since they are perfect, they achieve the bound (3.36) in Chapter 3.

Chapter 5

Capacity of steganographic channels

In this chapter we formalize the capacity of a steganographic scheme with respect to a detection function g, which partition the set $\mathcal{A}^n$ to permissible and non permissible set.

We use the following notation: random variable are denoted by capital letters (e.g. X), and their realizations by respective lower case letters (e.g. x). The domains over that random variables are denoted by script letters (e.g. $\mathcal{A}$). sequences of n random variables are denoted boldface symbols (e.g. $\mathbf{X} = (X_1, \ldots, X_n)$) which takes its values on the product set $\mathcal{A}^n$.

5.1 Permissible set.

The notion of permissible set is given by Harmsen and Pearlman (2005) in which they give the definition of secure steganographic capacity. We first define the detector function or steganalizer which classifies all output sequences into permissible and impermissible.

Definition 5.1. *A* steganalizer $g : \mathcal{A}^n \to \{0, 1\}$ *is a binary function that classifies the cover set* $\mathcal{A}^n$ *into two sets: containing steganographic information and not. That is for all* $\mathbf{y} \in \mathcal{A}^n$,

$$g(\mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{y} \text{ is steganographic}^1 \\ 0 & \text{otherwise.} \end{cases} \tag{5.1.1}$$

Let $\bar{d}_H : \mathcal{A}^n \times \mathcal{A}^n \to [0, 1]$ be the normalized Hamming distance on $\mathcal{A}^n$ such that, for all $\mathbf{x}, \mathbf{y} \in \mathcal{A}^n$,

$$\bar{d}_H(\mathbf{x}, \mathbf{y}) := \frac{1}{n} d_H(\mathbf{x}, \mathbf{y}) \tag{5.1.2}$$

where $d_H(\mathbf{x}, \mathbf{y})$ is the Hamming distance between $\mathbf{x}$ and $\mathbf{y}$ (see Definition 3.6).

We can define the steganalizer induced by $\bar{d}_H$.

Definition 5.2. *Let $D \in [0,1]$. The steganalizer induced by $\bar{d}_H$ is given such that for all $\mathbf{x}, \mathbf{y} \in \mathcal{A}$,*

$$g_H(\mathbf{y}) = \begin{cases} 0 & \text{if } \bar{d}_H(\mathbf{x},\mathbf{y}) \leq D \\ 1 & \text{otherwise.} \end{cases} \tag{5.1.3}$$

This detection function considers any sequence $\mathbf{y} \in \mathcal{A}^n$ at Hamming distance bigger than nD away from the cover-sequence $\mathbf{x} \in \mathcal{A}^n$ as steganographic.

A steganalyzer partitions $\mathcal{A}^n$ into permissible and impermissible where the two sets are defined as follows.

Definition 5.3. *For a given steganalizer g, the permissible set $\mathcal{P}_g \subseteq \mathcal{A}^n$ is the inverse image of 0 under g:*

$$\mathcal{P}_g := g^{-1}(\{0\}) = \{\mathbf{y} \in \mathcal{A}^n \mid g(\mathbf{y}) = 0\}. \tag{5.1.4}$$

Each element of the permissible set is classified as non-steganographic. The set of sequences that are classified as steganographic are impermissible for g.

Definition 5.4. *For a given steganalizer g, the impermissible set $\mathcal{I}_g \subseteq \mathcal{A}^n$ is the inverse image of 1:*

$$\mathcal{P}_g := g^{-1}(\{1\}) = \{\mathbf{y} \in \mathcal{A}^n \mid g(\mathbf{y}) = 1\}. \tag{5.1.5}$$

By definition, $\mathcal{P}_g$ and $\mathcal{I}_g$ partition $\mathcal{A}^n$. That is:

$$\mathcal{A}^n = \mathcal{P}_g \cup \mathcal{I}_g$$

and

$$\mathcal{P}_g \cap \mathcal{I}_g = \emptyset.$$

Lemma 5.5. *Let the detector function be g_H (see Definition 5.2) and $D \in [0,1]$. Then for any cover $\mathbf{x} \in \mathcal{A}^n$, the permissible and impermissible sets are given respectively by*

$$\mathcal{P}_{g_H}(\mathbf{x}) = B(\mathbf{x}, nD),$$

and

$$\mathcal{I}_{g_H}(\mathbf{x}) = {}^2\overline{B(\mathbf{x}, nD)}.$$

Proof. For any cover $\mathbf{x} \in \mathcal{A}^n$ and a real number $D \in [0,1]$, the inverse image of 0 is

$$\begin{aligned} g_H^{-1}(\{0\}) &:= \{\mathbf{y} \in \mathcal{A}^n | g_H(\mathbf{y}) = 0\} && \text{by definition of inverse image} \\ &= \{\mathbf{y} \in \mathcal{A}^n | \bar{d}_H(\mathbf{x},\mathbf{y}) \leq D\} && \text{by Equation 5.1.3} \\ &= \{\mathbf{y} \in \mathcal{A}^n | d_H(\mathbf{x},\mathbf{y}) \leq nD\} && \text{by Equation 5.1.2} \\ &= B(\mathbf{x}, nD). \end{aligned}$$

[2] $\overline{B(\mathbf{x}, nD)}$ denotes the complement set of $B(\mathbf{x}, nD)$.

Similarly, the inverse image of 1 is

$$\begin{aligned} g_H^{-1}(\{1\}) &:= \{\mathbf{y} \in \mathcal{A}^n | g_H(\mathbf{y}) = 1\} \\ &= \{\mathbf{y} \in \mathcal{A}^n | \bar{d}_H(\mathbf{x}, \mathbf{y}) > D\} \\ &= \{\mathbf{y} \in \mathcal{A}^n | d_H(\mathbf{x}, \mathbf{y}) > nD\} \\ &= \overline{B(\mathbf{x}, nD)} \end{aligned}$$

which is the complement of $B(\mathbf{x}, nD)$, hence satisfies

$$B(\mathbf{x}, nD) \cup \overline{B(\mathbf{x}, nD)} = \mathcal{A}^n$$

and

$$B(\mathbf{x}, nD) \cap \overline{B(\mathbf{x}, nD)} = \emptyset.$$

□

Given a cover-sequence $\mathbf{x} \in \mathcal{A}$ and an integer $d \geq 0$, for any $\mathbf{y} \in \mathcal{A}^n$, the steganalyzer g_H is equivalent to

$$g(\mathbf{y}) = \begin{cases} 0 & \text{if } d(\mathbf{x}, \mathbf{y}) \leq D \\ 1 & \text{otherwise} \end{cases} \tag{5.1.6}$$

Then the permissible set is the set of all sequences that are distance at most d away from the cover. That is

$$\begin{aligned} \mathcal{P}_g &= \{\mathbf{y} \in \mathcal{A} | d_H(\mathbf{x}, \mathbf{y}) \leq d\} \\ &:= B(\mathbf{x}, d). \end{aligned}$$

5.2 Steganographic capacity

We model the cover as a sequence $\mathbf{X} = (X_1, \ldots, X_n)$ of independent and identically distributed (i.i.d.) samples drawn from a probability mass function $p_X(x)$. The message $\mathbf{M}$ is to be embedded in $\mathbf{X}$ and transmitted. $\mathbf{M}$ is uniformly distributed over a message set $\mathcal{M}$. The embedding function produces a stego-sequence $\mathbf{Y}$ in order to transmit the message $\mathbf{M}$ reliably. The cover and stego-sequence are required to be close according to some distortion metric.

Definition 5.6. *A distortion function for the steganographer is a non negative function $d : \mathcal{A} \times \mathcal{A} \to [0, \infty)$, which can be extended to per-symbol distortions on n-tuples by*

$$d(\mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, y_i) \tag{5.2.1}$$

for $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$.

We assume without loss of generality that $d(x,y) \geq 0$, with equality if $x = y$ and $\max_{x,y} d(x,y) < \infty$. The classical distortion function is the Hamming distance.

Definition 5.7. *(Moulin and O'Sullivan, 2003) A steganographic scheme subject to distortion D is a triplet $(Emb, Ext, \mathcal{M})$ where*

- *$\mathcal{M}$ is the message set;*
- *$Emb : \mathcal{A}^n \times \mathcal{M} \to \mathcal{A}^n$ is the embedding function subject to distortion constraint D, i.e. for all $\mathbf{x} \in \mathcal{A}^n$ and $\mathbf{s} \in \mathcal{M}$*

$$d(\mathbf{x}, f_n(\mathbf{x}, \mathbf{s})) \leq D; \tag{5.2.2}$$

- *$Ext : \mathcal{A}^n \to \mathcal{M}$ is the extracting function.*

An (n, M, δ)-code consists of an embedding and extracting functions such that the encoder is capable of transferring one of the M secrets using a sequence of length n with a probability of detection less than δ.

Definition 5.8. *Given a steganalizer g, a sequence $\mathbf{y} \in \mathcal{A}^n$ is called δ-secure if*

$$Pr(g(\mathbf{y}) = 1) \leq \delta. \tag{5.2.3}$$

The set, $\mathcal{T}_\delta$, of all δ-secure sequences is

$$\mathcal{T}_\delta := \{\mathbf{Y} | P_{\mathbf{Y}}(\mathcal{I}_g) \leq \delta\}. \tag{5.2.4}$$

For $\delta = 0$, the set $\mathcal{T}_0$ is called the secure output set.

Lemma 5.9. *Given g, a secure output $\mathbf{Y} \in \mathcal{T}_0$ is permissible, i.e. $\mathbf{Y} \in \mathcal{P}_g$.*

Proof.

$$\begin{aligned} \mathcal{T}_0 &:= \{\mathbf{Y} | P_Y(\mathcal{I}_g) = 0\} \\ &= \{\mathbf{Y} | P_Y(\mathcal{P}_g) = 1\} \\ &= \{\mathbf{Y} | g(\mathbf{Y}) = 0\} \\ &= \{\mathbf{Y} \in \mathcal{P}_g\}. \end{aligned}$$

□

If the detection function is g_H (see Definition 5.2) and the cover set $\mathcal{A}^n$ is uniformly distributed, then for any cover $\mathbf{x} \in \mathcal{A}^n$ and any real $D \in [0,1]$

$$Pr(g(\mathbf{y}) = 1) = Pr(\mathbf{y} \in \overline{B(\mathbf{x}, nD)}). \tag{5.2.5}$$

Since

$$|B(\mathbf{x}, nD)| = \sum_{i=0}^{nD} \binom{n}{i} (|\mathcal{A}| - 1)^i, \tag{5.2.6}$$

then

$$Pr(\mathbf{y} \in \overline{B(\mathbf{x}, nD)}) = \frac{1}{|\mathcal{A}^n| - \sum_{i=0}^{nD} \binom{n}{i} (|\mathcal{A}| - 1)^i}. \tag{5.2.7}$$

The perfectly secure output set is

$$\begin{aligned} \mathcal{T}_0 &= \{\mathbf{y} \in \mathcal{A}^n | Pr(g(\mathbf{y}) = 1) = 0\} && (5.2.8) \\ &= \overline{B(\mathbf{x}, nD)}. && (5.2.9) \end{aligned}$$

For a given δ, we can derive the cardinality of the permissible (or impermissible) set. Moreover if the detector function is g_H (see Definition 5.2), then we can derive the volume of the ball and therefore the maximum D such that $B(\mathbf{x}, nD) \subseteq \mathcal{P}_{g_H}(\mathbf{x})$, for all cover $\mathbf{x} \in \mathcal{A}^n$.

The capacity of a steganographic scheme is the maximum of all achievable rate, where rate is defined as follows:

Definition 5.10. *The* rate *of the steganographic scheme is*

$$\alpha = \frac{1}{n} \log_2 |\mathcal{M}|. \tag{5.2.10}$$

Harmsen and Pearlman (2005) define secure steganographic capacity as follows:

Definition 5.11. *Given a steganalizer g, the secure capacity of a stego-scheme is given by*

$$C(g) := \max_{\mathbf{Y} \in \mathcal{T}_0} H(\mathbf{Y}). \tag{5.2.11}$$

Theorem 5.12. *For the staganalyzer g, the secure capacity is given by*

$$C(g) = \log_2 |\mathcal{P}_g|. \tag{5.2.12}$$

Proof. The maximum in Definition 5.11 is achieved if $\mathbf{Y}$ is uniformly distributed over the permissible set $\mathcal{P}_g$ (see Lemma 5.9). Thus the capacity satisfies

$$C(g) = \log_2 |\mathcal{P}_g|. \tag{5.2.13}$$

□

5.3 Examples

Now let us find the capacity of some steganalizers.

5.3.1 Typical set steganalizer

If the type stego-sequence is the same as the cover-sequence distribution, then it is considered non-steganographic.

Assume that P_X is the distribution over the finite alphabet $\mathcal{A}$. Let $\mathbf{x} \in \mathcal{A}^n$ be a sequence chosen by the detector to define a steganalizer which is defined as follows:

Definition 5.13. *The steganalizer specified by* $\mathbf{x}$ *for any sequence* $\mathbf{y} \in \mathcal{A}^n$ *is given as*

$$g_{\mathbf{x}}(\mathbf{y}) = \begin{cases} 1 & \text{if } P_{\mathbf{y}} \neq P_{\mathbf{x}} \\ 0 & \text{if } P_{\mathbf{y}} = P_{\mathbf{x}} \end{cases}$$

if $P_{\mathbf{y}}$ *is the type of the sequence* $\mathbf{y}$.

The permissible set for $g_{\mathbf{x}}$ is equal to the type class of $P_{\mathbf{x}}$, i.e.,

$$\begin{aligned} \mathcal{P}_{g_{\mathbf{x}}} &= T(P_{\mathbf{x}}) \\ &:= \{\mathbf{y} \in \mathcal{A}^n | P_{\mathbf{y}} = P_{\mathbf{x}}\}. \end{aligned}$$

Theorem 5.14. *The capacity of the typical set (see Appendix B) steganalizer specified by the sequence* $\mathbf{x} \in \mathcal{A}^n$ *is given by*

$$C(g_{\mathbf{x}}) = H(\mathbf{X}). \tag{5.3.1}$$

Proof. The steganographic capacity given the steganalizer $g_{\mathbf{x}}$ is

$$\begin{aligned} C(g_{\mathbf{x}}) &:= \log_2 |\mathcal{P}_{g_{\mathbf{x}}}| && \text{by Theorem 5.12} \\ &= \log_2 |T(P_{\mathbf{x}})| && \text{by definition of } g_x \\ &= \log_2 2^{nH(X)} \\ &= nH(X) \\ &= H(\mathbf{X}), \end{aligned}$$

where $\mathbf{X}$ is n i.i.d. realizations of the random variable X over $\mathcal{A}$. □

5.3.2 Bernouilli Hamming case

Here we consider $\mathcal{A} = \{0, 1\}$ and the distortion to be the Hamming distance. We assume that $Pp_X(1) = p_X(0) = 1/2$, i.e. X is here a Bernoulli$\left(\frac{1}{2}\right)$ source.

In this case, the capacity of a steganographic scheme with distortion constraint D, such that the steganalizer g_H is induced by Hamming distance, is given by Moulin and O'Sullivan (2003) as follows:

Lemma 5.15. *For Bernouilli*$\left(\frac{1}{2}\right)$*-Hamming with distortion constraint* D*, the hiding capacity is*

$$C(g_H, D) = \begin{cases} H(D) & 0 \leq D \leq \frac{1}{2} \\ 1 & \frac{1}{2} < D \end{cases} \tag{5.3.2}$$

where $H(D) = -D \log_2 D - (1 - D) \log_2 (1 - D)$.

From Lemma 5.15, a bound on the length of binary stego-schemes can be derived.

Theorem 5.16. *A stego-scheme from a linear $[n, k, \rho]_2$-covering code, such that $\frac{\rho}{n} \leq \frac{1}{2}$ must satisfy*

$$\frac{n-k}{n} \leq H\left(\frac{\rho}{n}\right).$$

Proof. By construction (see Chapter 4), a stego-scheme derived from a binary linear $[n, k, \rho]$-covering code has rate equal to $\frac{n-k}{n}$. By definition of capacity, $\frac{n-k}{n} \leq C(D)$, such that D is the maximum distortion, i.e. nD is the embedding radius which is equal to the covering radius of the code. Therefore we have,

$$\frac{n-k}{n} \leq C\left(\frac{\rho}{n}\right) = H\left(\frac{\rho}{n}\right).$$

□

For the stego-scheme arising from the binary Hamming $[2^r - 1, 2^r - 1 - r]$-code (see Example 4.15), the rate is $\frac{r}{2^r-1}$. Since the covering radius of Hamming codes is 1, then the capacity is $C\left(\frac{1}{2^r-1}\right) = H\left(\frac{1}{2^r-1}\right)$. Therefore we have $\frac{r}{2^r-1} \leq H\left(\frac{1}{2^r-1}\right)$. Actually, this bound is similar to the bound given in Theorem 3.41 of Chapter 3.

Chapter 6

Conclusion

The main goal of this thesis has been to reformalise the construction of embedding schemes with small embedding distortion in order to increase the embedding efficiency and decrease the chance of detection. It was firstly noticed by Crandall (1998) and Bierbrauer (1998) that linear codes can be transformed into a hiding scheme with small distortion. Since, researchers started to study the link between coding theory and steganography. Most results are on steganographic schemes based on binary error correcting codes. But in this work we have presented the general concept of the design of a stego-scheme with small embedding distortion.

We have seen that the construction of steganographic schemes with high embedding efficiency (or small embedding distortion) depends mostly on the extracting function $Ext : \mathcal{X} \to \mathcal{M}$. The latter is equivalent to an $|\mathcal{M}|$-partition of the cover set $\mathcal{X}$ and the embedding radius (or the maximum embedding distortion) of the scheme is the covering radius of the $|\mathcal{M}|$-partition. Therefore we need to choose an $|\mathcal{M}|$-partition with smallest covering radius. The embedding function has also its role to improve the efficiency of the stego-scheme. For the stego and the cover to be as close as possible, the embedding function, Emb, must satisfy the maximum likelihood decoding problem. It finds a nearest element to the cover in the corresponding subset $\mathcal{X}_{\mathbf{s}} \subseteq \mathcal{X}$, where $\mathcal{X}_{\mathbf{s}} = Ext^{-1}(\{\mathbf{s}\})$.

Note that the subsets in an $|\mathcal{M}|$-partition of $\mathcal{X}$ need not have the same cardinality. All that we required is for the covering radius to be small.

The relationship between coding theory and steganopgraphy has been given in Chapter 3 as a specification of the previous method, where the covers are strings of symbols (of length n) from a finite alphabet $\mathcal{A}$ that has a structure of an Abelian group. The cover set is the set of all n-tuples $\mathcal{A}^n$ and the set of secrets is $\mathcal{M}$. Then we consider the $|\mathcal{M}|$-partition of $\mathcal{A}^n$ to be the quotient space $\mathcal{A}^n/\mathcal{C}$ ($\mathcal{C}$ is a subgroup of $\mathcal{A}^n$), where $\mathcal{C}$ is a block code defined on $\mathcal{A}$. The advantage of this construction is that the decoding procedure on each coset (elements of the $\mathcal{A}^n/\mathcal{C}$) derives only from the decoding map of the code $\mathcal{C}$. We can derive some bounds on the performance of stego-schemes from codes (or

code-based stego-scheme) by applying some classic bounds in coding theory.

Crandall's discovery, matrix embedding, is presented in Chapter 4, which uses linear codes to embed data. In this case secrets and covers are strings from the finite field $\mathbb{F}_q$ (q a power of a prime) but with different length, say k and n respectively. The secret is extracted as syndrome of the received stego-text, with respect to a parity check matrix of the linear code. We have derived in this method the correspondence between the parameters of the stego-schemes and linear codes.

The last chapter focuses on the capacity of a stego-scheme, where we gave some examples. We have seen in this part that even though the capacity of a steganographic source is not easy to compute, the capacity of the Bernoulli Hamming source and the upper bound of the rate in Chapter 3 coincide. So it might be possible that bounds provided by coding theoretic methods answer the problem of finding the capacity for a general source.

We justified in the concluding that coding theory is useful in steganogarphy more generally than previously studied. In particular it has enabled us to address and answer questions such as the maximum number of changes needed to embed a secret, and the maximum embeddable secret length.

There are still many other problems remaining, such as the bounds on the length of the stego-scheme, which is defined by the length of the cover, the construction of fast embedding algorithms and the construction of schemes that approach the hiding capacity.

Appendices

Appendix A

Entropy function

Definition A.1. *Let $q > 1$ be an integer and $0 \leq p \leq 1$ be a real. Then the q-ary entropy function is defined as follows:*

$$H_q(p) = p\log_q(q-1) - p\log_q(p) - (1-p)\log_q(1-p).$$

If $q = 2$, the binary entropy function is given by

$$H(p) = -p\log_2 p - (1-p)\log_2(1-p).$$

The function H_q is continuous and increasing in the interval $[0, 1-1/q]$ with the convention: $H_q(0) = 0$ and $H_q(1-1/q) = 1$. Figure A.1 gives a representation of the $H_q(.)$ for the some few values of q. It shows that the binary entropy function is symmetric around the line $p = 1/2$: $H(1-p) = H(p)$.

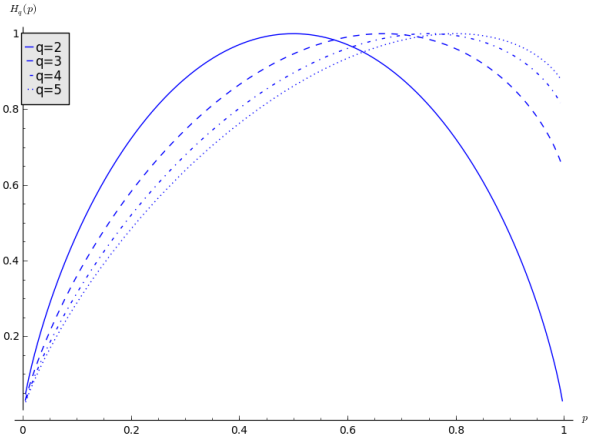

Figure A.1: A plot of H_q for $q = 2, 3, 4, 5$.

Appendix B

Typical set

The main result of the method of the type given in (Cover and Thomas, 2012) states that the outputs of n independent and identically distributed (i.i.d.) realizations of random variables X can be partitioned into two sets, typical and non-typical, where the probability of the typical set is nearly 1, moreover the typical set is equally distributed and has nearly $2^{nH(X)}$ elements, called typical sequences. That result can be applied to our concept in such a way that the detection function classifies any sequence that is outside the typical set as steganographic.

Let $\mathcal{X}$ be a finite alphabet and $n \in \mathbb{N}$. Let $\mathbf{x}^n$ be a sequence of n symbols from a finite alphabet $\mathcal{X}$.

Definition B.1. *The type of a sequence $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ is the probability mass function $P_{\mathbf{x}}$ defined as the relative proportion of occurrences of each symbol of $\mathcal{X}$. That is, for all $a \in \mathcal{X}$,*

$$P_{\mathbf{x}}(a) = \frac{N_{\mathbf{x}}(a)}{n} \tag{B.0.1}$$

where $N_{\mathbf{x}}(a) = |\{i|x_i = a\}|$ is the number of times the symbol a occurs in the sequence $\mathbf{x} \in \mathcal{X}^n$.

Example B.2. *Let $\mathcal{X} = \{0, 1\}$ and $n = 8$. The type of $\mathbf{x} = (1, 0, 0, 0, 1, 0, 1, 0)$ is $(P_{\mathbf{x}}(0) = \frac{5}{8}, P_{\mathbf{x}}(1) = \frac{3}{8})$.*

Definition B.3. *The set of all possible types of sequences $\mathbf{x} \in \mathcal{X}^n$ is denoted by $\mathcal{P}_n(\mathcal{X})$.*

Similarly we define the joint type of two sequences as follows.

Definition B.4. *The joint type of two sequences $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ and $\mathbf{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$ is the probability distribution $P_{\mathbf{xy}} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ defined as the relative proportion of occurrences of each pair $(a, b) \in \mathcal{X} \times \mathcal{Y}$ among the (x_i, y_i) for all $i \in [1, n]$.*

Example B.5. *If $\mathcal{X} := \mathbb{B}$, then the set of types of pairs of sequences of length n is*

$$\mathcal{P}_n = \left\{ \left(\frac{i}{n}, \frac{n-i}{n} \right) | 0 \leq i \leq n \right\}. \tag{B.0.2}$$

Definition B.6. *Let $P \in \mathcal{P}_n$. Then the set of sequences of length n and type P is called the type class of P, denoted $T(P)$, i.e.*

$$T(P) = \{\mathbf{x} \in \mathcal{X}^n | P_{\mathbf{x}} = P\}. \tag{B.0.3}$$

Example B.7. *Let $\mathcal{X} = \{0, 1\}$ and $\mathbf{x} = (1, 0, 0, 1, 0)$. The type class of $P_{\mathbf{x}}$ is the set of all sequences of length 5 with three 0's and two 1's. There are 10 of them*

$$T(P_{\mathbf{x}}) = \quad \{\mathbf{x} \in \mathcal{X}^5 | N_{\mathbf{x}(1)=2}\}.$$

Now assume that the sequence $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is drawn i.i.d. according to a distribution $Q(\mathbf{x})$. Then all sequences of the same type have the same probability.

Theorem B.8. *If $X_1, X_2, \ldots, X_n$ are drawn i.i.d. according to $Q(x)$, then the probability of $\mathbf{x}$ depends only on its type and is given by*

$$Q^n(\mathbf{x}) = 2^{n(D(P_{\mathbf{x}}||Q) - P_{\mathbf{x}})} \tag{B.0.4}$$

where $D(P_{\mathbf{x}}||Q) = \sum_{\mathbf{x} \in \mathcal{X}} P_{\mathbf{x}}(\mathbf{y}) \log \frac{P_{\mathbf{x}}(\mathbf{y})}{Q(\mathbf{y})}$ is the relative entropy between $P_{\mathbf{x}}$ and Q.

Proof.

$$\begin{aligned}
Q^n(\mathbf{x}) &= \prod_{i=n}^{n} Q(x_i) \\
&= \prod_{a \in \mathcal{X}} Q(a)^{N_{\mathbf{x}}(a)} \\
&= \prod_{a \in \mathcal{X}} n P_{\mathbf{x}}(a) \\
&= \prod_{a \in \mathcal{X}} 2^{n P_{\mathbf{x}}(a) \log Q(a)} \\
&= \prod_{a \in \mathcal{X}} 2^{n(P_{\mathbf{x}}(a) \log Q(a) - P_{\mathbf{x}}(a) \log P_{\mathbf{x}}(a) + P_{\mathbf{x}}(a) \log P_{\mathbf{x}}(a))} \\
&= 2^{n \prod_{a \in \mathcal{X}} (-P_{\mathbf{x}}(a) \log \frac{P_{\mathbf{x}}(a)}{Q(a)} + P_{\mathbf{x}}(a) \log P_{\mathbf{x}}(a))} \\
&= 2^{n(D(P_{\mathbf{x}}||Q) - H(P_{\mathbf{x}}))}.
\end{aligned}$$

□

Corollary B.9. *If* $\mathbf{x}$ *is in the type class of* Q*, then*

$$Q^n(\mathbf{x}) = 2^{-nH(Q)}. \tag{B.0.5}$$

Proof. If $\mathbf{x} \in T(Q)$, then $P_{\mathbf{x}} = Q$. Substituting to (B.8) we obtain the result. □

The size $|T(P)|$ of a type class is the number of ways of arranging $nP(a_1), \ldots, nP(a_{|\mathcal{X}|})$ objects in a sequence, which is the multinomial coefficient

$$|T(P)| = \binom{n}{nP(a_1), \ldots, nP(a_{|\mathcal{X}|})}. \tag{B.0.6}$$

Theorem B.10. *For any type* $P \in \mathcal{P}_n$*, the size of the type class of* P *can be estimated as follows:*

$$|T(P)| \simeq 2^{nH(P)}. \tag{B.0.7}$$

Lemma B.11. *For* $\mathcal{X} = \mathbb{B}$*, the type is defined by the weight of a sequence, and the size of the type class is therefore* $\binom{n}{k}$ *and it is bounded as follows*

$$\binom{n}{k} \leq 2^{nH\left(\frac{k}{n}\right)}. \tag{B.0.8}$$

Proof. The binomial formula says that

$$\sum_{k=0}^{n} \binom{n}{k} \left(\frac{k}{n}\right)^k \left(1 - \left(\frac{k}{n}\right)\right)^{n-k} = 1. \tag{B.0.9}$$

Therefore, by taking the k^{th} term, we get

$$\begin{aligned} 1 &\geq \binom{n}{k} \left(\frac{k}{n}\right)^k \left(1 - \left(\frac{k}{n}\right)\right)^{n-k} \\ &= \binom{n}{k} 2^{k \log \frac{k}{n} + (n-k) \log \frac{n-k}{n}} \\ &= \binom{n}{k} 2^{n\left(\frac{k}{n} \log \frac{k}{n} + \frac{n-k}{n} \log \frac{n-k}{n}\right)} \\ &= \binom{n}{k} 2^{-nH\left(\frac{k}{n}\right)}. \end{aligned}$$

Hence

$$\binom{n}{k} \leq 2^{nH\left(\frac{k}{n}\right)}. \tag{B.0.10}$$

□

Theorem B.12. *For any type* $P \in \mathcal{P}_n$ *and any distribution* Q*, the probability of the type class* $T(P)$ *under* Q *satisfies*

$$Q(T(P)) \leq 2^{nD(P||Q)}. \tag{B.0.11}$$

Proof. We have

$$\begin{aligned} Q(T(P)) &= \sum_{\mathbf{x}\in T(P)} Q(\mathbf{x}) \\ &= \sum_{\mathbf{x}\in T(P)} 2^{-n(D(P||Q)+H(P))} \\ &= |T(P)|2^{-n(D(P||Q)+H(P))}. \end{aligned}$$

Using the bounds on $T(P)$, we have

$$Q(T(P)) \leq 2^{nD(P||Q)}. \tag{B.0.12}$$

□

Given any $\epsilon > 0$, we can define a typical set T_Q^ϵ for the distribution Q as

$$T_Q^\epsilon = \{\mathbf{x} \in \mathcal{X}^n | D(P_\mathbf{x}||Q) \leq \epsilon\}. \tag{B.0.13}$$

Then the probability that $\mathbf{x}$ is not typical is

$$1 - Q(T_Q^\epsilon) = \sum_{P:D(P||Q)>\epsilon} Q(T(P)) \tag{B.0.14}$$

$$\leq \sum_{P:D(P||Q)>\epsilon} 2^{-n(D(P||Q)+H(P))} \tag{B.0.15}$$

$$\leq \sum_{P:D(P||Q)>\epsilon} 2^{-n\epsilon} \tag{B.0.16}$$

$$\leq (n+1)^{|\mathcal{X}|}2^{-n\epsilon} \tag{B.0.17}$$

$$= 2^{-n\left(\epsilon - |\mathcal{X}|\frac{\log(n-1)}{n}\right)}, \tag{B.0.18}$$

which goes to 0 as n tends to ∞, and then the probability of the typical set goes to 1.

Theorem B.13. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. according to $P(x)$. Then*

$$Pr\{D(P_\mathbf{x}||P) > \epsilon\} \leq 2^{-n\left(\epsilon - |\mathcal{X}|\frac{\log(n-1)}{n}\right)}. \tag{B.0.19}$$

Thus $D(P_\mathbf{x}||P) \longrightarrow 0$ with probability 1.

Proof. Summing over n, we find that

$$\sum_{n=1}^{\infty} Pr\{D(P_\mathbf{x}||P) > \epsilon\} < \infty. \tag{B.0.20}$$

Thus the expected number of occurrences of the even $D(P_\mathbf{x}||P) > \epsilon$ for all n finite, which implies that the actual number of such occurences is also finite with probability 1. Hence $D(P_\mathbf{x}||P) \longrightarrow 0$ with probability 1. □

A stronger version of typicality is given as follows.

Definition B.14. *The strongly typical set A_ϵ is the set of sequences in $\mathcal{X}^n$ for which the frequencies are close to the true values, i.e.,*

$$A_\epsilon = \left\{ \mathbf{x} \in \mathcal{X}^n | \left| \frac{1}{n} N_{\mathbf{x}}(a) - P(a) \right| < \frac{\epsilon}{|\mathcal{X}|}, \text{ for all } a \in \mathcal{X} \right\}. \tag{B.0.21}$$

The typical set consists of sequences whose type does not differ from the true probabilities by more than $\frac{\epsilon}{|\mathcal{X}|}$ in any component. By the strong law of large numbers it follows that the probability of the strongly typical set goes to 1 as n increases.

List of References

Barbier, M. (2010). New set of codes for the maximum-likelihood decoding problem. *arXiv preprint arXiv:1011.2834*.

Bierbrauer, J. (1998). On crandall's problem. *Personal communication available from* `http://www.ws.binghamton.edu/fridrich/covcodes.pdf`.

Cachin, C. (1998). An information-theoretic model for steganography. In: *Information Hiding*, pp. 306–318. Springer.

Cover, T.M. and Thomas, J.A. (2012). *Elements of Information Theory*. John Wiley & Sons.

Crandall, R. (1998). Some notes on steganography. *Posted on steganography mailing list*.

Engle, S. (2003). Current state of steganography: Uses, limits, & implications. *Class Paper*, , no. CS 235 Computer Security.

Forney, G. (1992 Nov). On the hamming distance properties of group codes. *Information Theory, IEEE Transactions on*, vol. 38, no. 6, pp. 1797–1801. ISSN 0018-9448.

Fridrich, J. (2009). *Steganography in digital media: principles, algorithms, and applications*. Cambridge University Press.

Fridrich, J., Lisoněk, P. and Soukal, D. (2007*a*). On steganographic embedding efficiency. In: *Information Hiding*, pp. 282–296. Springer.

Fridrich, J., Pevnỳ, T. and Kodovskỳ, J. (2007*b*). Statistically undetectable jpeg steganography: dead ends challenges, and opportunities. In: *Proceedings of the 9th workshop on Multimedia & security*, pp. 3–14. ACM.

Fridrich, J. and Soukal, D. (2006). Matrix embedding for large payloads. In: *Electronic Imaging 2006*, pp. 60721W–60721W. International Society for Optics and Photonics.

Galand, F. and Kabatiansky, G. (2003*a*). Information hiding by coverings. In: *Information Theory Workshop, 2003. Proceedings. 2003 IEEE*, pp. 151–154. IEEE.

Galand, F. and Kabatiansky, G. (2003*b*). Information hiding by coverings. In: *Information Theory Workshop, 2003. Proceedings. 2003 IEEE*, pp. 151–154. IEEE.

Harmsen, J.J. and Pearlman, W.A. (2005). Capacity of steganographic channels. In: *Proceedings of the 7th workshop on Multimedia and security*, pp. 11–24. ACM.

Ker, A.D. (2008). Steganographic strategies for a square distortion function. In: *Electronic Imaging 2008*, pp. 681904–681904. International Society for Optics and Photonics.

Ker, A.D., Pevnỳ, T., Kodovskỳ, J. and Fridrich, J. (2008). The square root law of steganographic capacity. In: *Proceedings of the 10th ACM workshop on Multimedia and security*, pp. 107–116. ACM.

MacWilliams, F.J. and Sloane, N.J.A. (1977). *The theory of error-correcting codes*, vol. 16. Elsevier.

Moulin, P. and O'Sullivan, J.A. (2003). Information-theoretic analysis of information hiding. *Information Theory, IEEE Transactions on*, vol. 49, no. 3, pp. 563–593.

Moulin, P. and Wang, Y. (2004). New results on steganographic capacity. In: *Proc. CISS Conference*.

Munuera, C. (2012). Steganography from a coding theory point of view. *Algebraic geometry modeling in information theory. Edited by E. Martinez-Moro. World Scientific*.

Munuera, C. and Barbier, M. (2011). Wet paper codes and the dual distance in steganography. *arXiv preprint arXiv:1104.1970*.

Ralaivaosaona, T.F. (May 2013). *Steganography*. Master's thesis, The African Institute for Mathematical Sciences. `http://docs.google.com/viewer?a=v&pid=sites&srcid=YWltcy5hYy56YXxhcmNoaXZlfGd4OjcxNDcwMTNkYWIzNmRmZTU`.

Schneier, B. *et al.* (1996). Applied cryptography: protocols, algorithms, and source code in c. *John Wiley & Sons, Inc*, vol. 2, p. 9.

Simmons, G.J. (1983). The prisoners' problem and the subliminal channel. In: *Advances in Cryptology: Proceedings of CRYPTO '83*, pp. 51–67. Plenum.

Smith, G. (2011 Sept). Quantifying information flow using min-entropy. In: *Quantitative Evaluation of Systems (QEST), 2011 Eighth International Conference on*, pp. 159–167.

Tietäväinen, A. (1973). On the nonexistence of perfect codes over finite fields. *SIAM Journal on Applied Mathematics*, vol. 24, no. 1, pp. 88–96.

Westfeld, A. (2001). F5 a steganographic algorithm. In: *Information Hiding*, pp. 289–302. Springer.

Zhang, W. and Li, S. (2005). Steganographic codes–a new problem of coding theory. *arXiv preprint cs/0505072*.

Zöllner, J., Federrath, H., Klimant, H., Pfitzmann, A., Piotraschke, R., Westfeld, A., Wicke, G. and Wolf, G. (1998). Modeling the security of steganographic systems. In: *Information Hiding*, pp. 344–354. Springer.